



**ANALISIS SENTIMEN UNTUK PEMERINGKATAN POPULARITAS
SITUS BELANJA ONLINE DI INDONESIA MENGGUNAKAN METODE
NAIVE BAYES (STUDI KASUS DATA SEKUNDER)**

TUGAS AKHIR

**Program Studi
S1 Sistem Informasi**

**INSTITUT BISNIS
DAN INFORMATIKA**

stikom
SURABAYA

Oleh :

FIRRARI KUSUMA WARDANI

15410100048

FAKULTAS TEKNOLOGI DAN INFORMATIKA

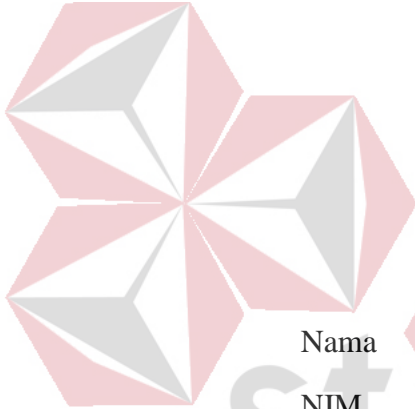
INSTITUT BISNIS DAN INFORMATIKA STIKOM SURABAYA

2019

**ANALISIS SENTIMEN UNTUK PEMERINGKATAN
POPULARITAS SITUS BELANJA ONLINE DI INDONESIA
MENGUNAKAN METODE NAIVE BAYES (STUDI KASUS
DATA SEKUNDER)**

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk menyelesaikan Program Sarjana
Sistem Informasi**



Disusun Oleh :

Nama : FIRRARI KUSUMA WARDANI
NIM : 15410100048
Program : S1 (Strata Satu)
Jurusan : Sistem Informasi

**FAKULTAS TEKNOLOGI DAN INFORMATIKA
INSTITUT BISNIS DAN INFORMATIKA STIKOM SURABAYA**

2019

TUGAS AKHIR

ANALISIS SENTIMEN UNTUK PEMERINGKATAN POPULARITAS SITUS BELANJA ONLINE DI INDONESIA MENGGUNAKAN METODE NAIVE BAYES (STUDI KASUS DATA SEKUNDER)

Dipersiapkan dan disusun oleh

FIRRARI KUSUMA WARDANI
NIM : 15410100048


Telah diperiksa, diuji dan disetujui oleh Dewan Penguji

Pada : 04 FEBRUARI 2019


SUSUNAN DEWAN PENGUJI

Pembimbing

- I. **Valentinus Roby Hananto, S.Kom., M.Sc.,**
NIDN. 0715028903

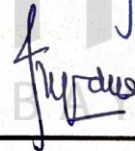
 26/2/19

- II. **Vivine Nurcahyawati, M.Kom.,**
NIDN. 0723018101

 26/02/19

Penguji

- I. **Dr. M. J. Dewiyani Sunarto**
NIDN. 0725076301



Tugas Akhir ini telah diterima sebagai salah satu persyaratan
Untuk memperoleh gelar sarjana



FAKULTAS TEKNOLOGI
DAN INFORMATIKA

stikom
SURABAYA

Dr. Jusak

Dekan Fakultas Teknologi dan Informatika

26/19
2

INSTITUT BISNIS DAN INFORMATIKA STIKOM SURABAYA

PERNYATAAN

PERSETUJUAN PUBLIKASI DAN KEASLIAN KARYA ILMIAH

Sebagai mahasiswa Institut Bisnis dan Informatika Stikom Surabaya, saya :

Nama : Firrari Kusuma Wardani

NIM : 15410100048

Program Studi : S1 Sistem Informasi

Fakultas : Fakultas Teknologi dan Informatika

Jenis Karya : Tugas Akhir

Judul Karya : **ANALISIS SENTIMEN UNTUK PEMERINGKATAN POPULARITAS SITUS BELANJA ONLINE DI INDONESIA MENGGUNAKAN METODE NAIVE BAYES (STUDI KASUS DATA SEKUNDER)**

Menyatakan dengan sesungguhnya bahwa:

1. Demi pengembangan Ilmu Pengetahuan, Teknologi dan Seni, saya menyetujui memberikan kepada Institut Bisnis dan Informatika Stikom Surabaya Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalti Free Right*) atas seluruh isi/ sebagian karya ilmiah saya tersebut di atas untuk disimpan, dialihmediakan dan dikelola dalam bentuk pangkalan data (*database*) untuk selanjutnya didistribusikan atau dipublikasikan demi kepentingan akademis dengan tetap mencantumkan nama saya sebagai penulis atau pencipta dan sebagai pemilik Hak Cipta
2. Karya tersebut di atas adalah karya asli saya, bukan plagiat baik sebagian maupun keseluruhan. Kutipan, karya atau pendapat orang lain yang ada dalam karya ilmiah ini adalah semata hanya rujukan yang dicantumkan dalam Daftar Pustaka saya
3. Apabila dikemudian hari ditemukan dan terbukti terdapat tindakan plagiat pada karya ilmiah ini, maka saya bersedia untuk menerima pencabutan terhadap gelar kesarjanaan yang telah diberikan kepada saya.

Demikian surat pernyataan ini saya buat dengan sebenarnya.

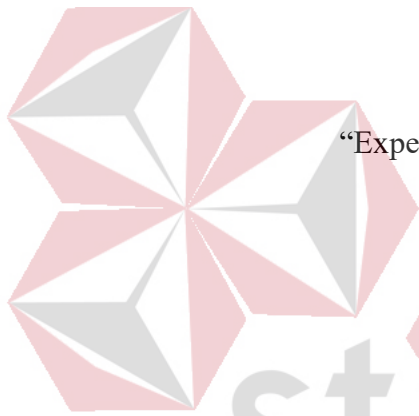
Surabaya, 14 Februari 2019



Firrari Kusuma Wardani

NIM : 15410100048

LEMBAR MOTTO



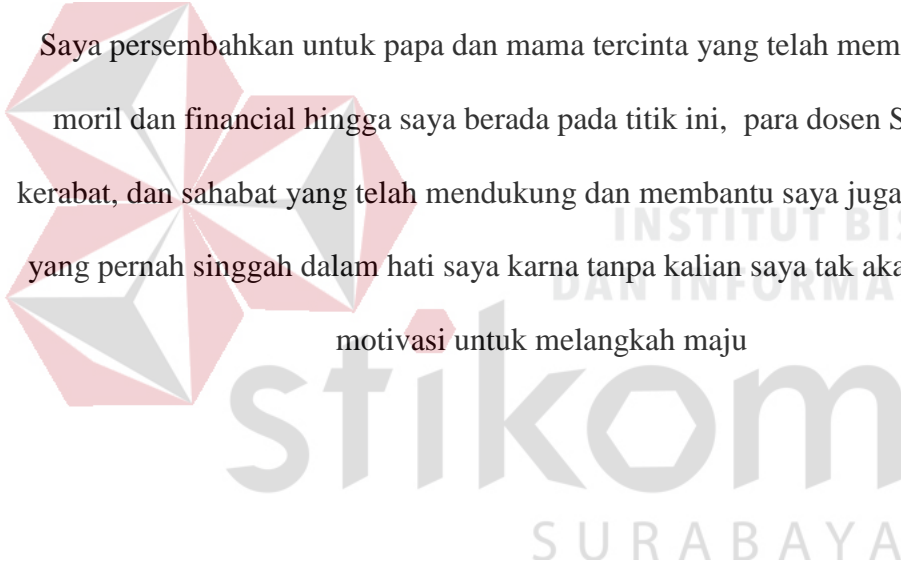
“Experience is the best teacher”

INSTITUT BISNIS
DAN INFORMATIKA

stikom
SURABAYA

LEMBAR PERSEMBAHAN

Saya persembahkan untuk papa dan mama tercinta yang telah memantu secara moril dan financial hingga saya berada pada titik ini, para dosen S1 SI, para kerabat, dan sahabat yang telah mendukung dan membantu saya juga orang-orang yang pernah singgah dalam hati saya karna tanpa kalian saya tak akan mendapat motivasi untuk melangkah maju



ABSTRAK

Lazada, Bukalapak dan Blibli.com adalah perusahaan yang terlibat dalam pasar dan *mal online*. Untuk menanggapi pendapat, kritik dan keluhan, Lazada, Bukalapak dan Blibli.com terdapat suatu *platform website* yang mengulas layanan dan jasa yang ditawarkan oleh ketiga perusahaan tersebut. *Platform website* tersebut bernama priceprice.com. Jenis komentarnya adalah teks tidak terstruktur dalam jumlah besar. Kondisi ini dapat menyebabkan perusahaan belanja *online* melewatkan informasi berguna dari koleksi dokumen teks.

Analisis setiment ditetapkan dalam studi kasus ini untuk melakukan pemeringkatan popularitas situs belanja *online* di Indonesia. Pada penelitian ini hanya membahas tiga *online Shop*, yakni Lazada, Bukalapak dan Blibli.com. Studi kasus ini menggunakan metode *Naïve Bayes Classifier* (NBC) dengan pengujian *10-fold cross validation* untuk memastikan keakuratan data. Metode NBC digunakan karena lebih cocok dalam pengoperasian analisis sentiment pada penelitian ini dibandingkan dengan metode lainnya.

Pada penelitian ini Lazada merupakan *online shop* yang paling populer dibandingkan dengan Bukalapak dan Blibli.com. Berdasarkan perhitungan pemeringkatan popularitasnya didapatkan untuk *online shop* Lazada memiliki tingkat popularitas 95,9%, Bukalapak 93,7% dan Blibli.com 89,3%. Dengan uji validitas metode NBC menggunakan *10-fold cross validation* didapatkan akurasi untuk Lazada 96,5%, Bukalapak 93,9% dan Blibli.com 90,3%.

Kata kunci : *Lazada, bukalapak, blibli.com, naive bayes, sentiment analysis.*

KATA PENGANTAR

Bismillahirrahmanirrahim, Assalamu‘alaikum wr. wb. Alhamdulillah segala puji syukur kami panjatkan kehadirat Allah SWT, karena atas limpahan rahmat dan hidayah-Nya penulis dapat menyelesaikan Laporan Tugas Akhir dengan judul ***”Analisis Sentimen Untuk Pemeringkatan Popularitas Situs Belanja Online Di Indonesia Menggunakan Metode Naive Bayes (Studi Kasus Data Sekunder)”*** dengan baik dan lancar.

Laporan ini disusun sebagai bukti bahwa Penulis telah menyelesaikan Tugas Akhir. Adapun tujuan utama pelaksanaan tugas akhir ini untuk memenuhi salah satu syarat menyelesaikan program Sarjana Komputer, selain itu ilmu yang penulis dapat di perkuliahan dapat diimplementasikan dalam dunia kerja.

Penyelesaian tugas akhir tidak lepas dari bantuan beberapa pihak yang benar-benar memberikan dukungan serta masukan kepada Penulis. Oleh karena itu pada kesempatan kali ini Penulis mengucapkan banyak terima kasih kepada:

1. Bapak, Ibu dan Keluarga atas doa serta dukungan yang telah diberikan kepada Penulis.
2. Ibu Vivine Nurcahyawati, M.Kom., dan Bapak Valetinus Roby Hananto, S.Kom., M.Sc., selaku Dosen Pembimbing yang telah mendukung serta memberikan kepercayaan penuh kepada Penulis dalam menyelesaikan tugas akhir.
3. Ibu Dr. M. J. Dewiyani Sunarto selaku dosen penguji.
4. Sahabat yang selalu memberikan motivasi dan dorongan setiap hari. Diska Febria Utami, Nadya Olievia Hapshari Putri, Agustina Tri Wahyuningsih,

Elok Nurjannah, Surya Dana Arum dan untuk kekasih hati tercinta Bayu Sapta Angkasa Putra yang selalu setia mendampingi penulis dalam mengerjakan dan menyelesaikan tugas akhir ini.

Penulis menyadari bahwa laporan ini masih jauh dari kata sempurna, baik dari materi maupun teknik penyajian. Untuk itu Penulis mengharapkan saran dan kritik yang membangun dari Pembaca demi penyempurnaan dalam menyelesaikan tugas-tugas lainnya.

Wassalamu'alaikum wr. wb.

Surabaya, Februari 2019

Penulis



INSTITUT BISNIS
DAN INFORMATIKA

FIRRARI KUSUMA WARDANI

stikom
SURABAYA

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN	iv
LEMBAR PERSETUJUAN PUBLIKASI.....	v
LEMBAR MOTTO	vi
LEMBAR PERNYATAAN	vii
LEMBAR MOTTO	viii
LEMBAR PERSEMBAHAN	ix
ABSTRAK	viii
KATA PENGANTAR	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah	6
1.4 Tujuan	6
1.5 Manfaat.....	6

BAB II KAJIAN PUSTAKA	7
2.1 E-Commerce	7
2.2 Analisis	7
2.3 Sentiment Analisis dan Text Mining.....	8
2.4 Naive Bayes Classifier.....	12
2.5 Lazada	15
2.6 Bukalapak.....	17
2.7 Blibi.com.....	18
2.8 Rapid Miner.....	20
2.9 Penelitian Sebelumnya	20
BAB III METODOLOGI PENELITIAN.....	22
3.1 Tahapan Awal Penelitian.....	23
3.1.1.Studi Literatur	23
3.1.2 Teknik pengumpulan Data	23
3.2 Tahap Analisis Data	29
3.2.1 Input Data	30
3.2.2 Data Preprocessing	30
3.2.3 Klasifikasi Naive Bayes	32
3.2.4 Analisis Validitas Data	33
3.2.5 Perbandingan Rating Secara Umum	37
3.2.6 Visualisasi Data dengan Word Cloud	38

BAB IV HASIL DAN PEMBAHASAN	40
4.1 Karakteristik Data dan Input Data.....	40
4.2 Data <i>Preprocessing</i>	46
4.3 Klasifikasi Naive Bayes (NBC)	61
4.4 Uji Validitas Data	69
4.5 Rating secara umum	72
4.6 Visualisasi data dengan Wordcloud	75
BAB V PENUTUP	78
5.1 Kesimpulan.....	78
5.2 Saran	79
DAFTAR PUSTAKA	81



INSTITUT BISNIS
DAN INFORMATIKA
stikom
SURABAYA

DAFTAR GAMBAR

	Halaman
Gambar 1.1 Pengguna Internet di dunia selama kurun waktu 6 tahun terakhir.	1
Gambar 1.2 Situs online di Indonesia dengan 10 peringkat teratas.	2
Gambar 2.1 Tahap Tokenizing.....	10
Gambar 2.2 Tahap Filtering	11
Gambar 2.3 Tahap Stemming	12
Gambar 2.4 Ilustrasi 4-fold cross-validation.....	14
Gambar 2.5 Logo Lazada.....	15
Gambar 2.6 Logo Bukalapak	17
Gambar 2.7 Logo Blibli.com	18
Gambar 3.1 Diagram metode penelitian	22
Gambar 3.2 Bagan Tahapan Awal	23
Gambar 3.3 Website priceprice komentar pengguna	25
Gambar 3.4 Diagram alir penelitian.....	28
Gambar 3.5 Diagram alir nbc.....	29
Gambar 3.6 Teknik Analisis Data.....	29
Gambar 3.7 Rapidminer input data.	30
Gambar 3.8 Data preprocessing.	30
Gambar 3.9 Tahapan data preprocessing.	31
Gambar 3.10 Memasukkan kamus Liu's word list.	31
Gambar 3.11 naive bayes read excel.....	32
Gambar 3.12 naive bayes sselect atributte.	32
Gambar 3.13 naive bayes classification	32

Gambar 3.14 Proses Training.....	34
Gambar 3.15 Hubungkan operator cross validation.....	34
Gambar 3.16 Sub-proses training dan testing.	35
Gambar 3.17 Apply model sub-proses testing.	35
Gambar 3.18 Performance testing.....	35
Gambar 3.19 Hubungkan output.	36
Gambar 3.20 Klik link proses.	36
Gambar 3.21 Hubungkan port hijau dan kuning.	37
Gambar 3.22 Tagcrowd visualisasi wordcloud.....	38
Gambar 3.23 Tagcrowd Upload File.....	39
Gambar 4.1 Data komentar priceprice tahun 2017 – 2018.	40
Gambar 4.2 Perkembangan Komentar Blibli.com.....	41
Gambar 4.3 Prosentase sentimen positif dan negatif Blibli.com.....	42
Gambar 4.4 Perkembangan Komentar Lazada.....	43
Gambar 4.5 Prosentase sentimen positif dan negatif Lazada.....	43
Gambar 4.6 Perkembangan Komentar Lazada.....	44
Gambar 4.7 Prosentase sentimen positif dan negatif Lazada.....	45
Gambar 4.8 Frekuensi data komentar lazada	55
Gambar 4.9 Frekuensi data komentar bukalapak.....	56
Gambar 4.10 Frekuensi data komentar blibli.....	57
Gambar 4.11 Visualisasi Wordcloud Lazada.....	76
Gambar 4.12 Visualisasi Wordcloud Bukalapak.	76
Gambar 4.13 Visualisasi Wordcloud Blibli.	77

DAFTAR TABEL

	Halaman
Tabel 1.1 Kelebihan dan kekurangan (Dewantoro,2016).	3
Tabel 2.1 Penelitian sebelumnya.....	20
Tabel 3.1 Merubah kata menjadi baku.	26
Tabel 3.2 Variabel penelitian	27
Tabel 3.3 perbandingan rating secara umum	38
Tabel 4.1 Keseluruhan Sentimen Positif dan Negatif.	46
Tabel 4.2 Proses case folding.....	46
Tabel 4.3 Proses Tokenizing.....	48
Tabel 4.4 Proses Filtering.	49
Tabel 4.5 Proses Stemming.....	50
Tabel 4.6 Frekuensi data setelah proses stemming Lazada.	52
Tabel 4.7 Frekuensi data setelah proses stemming Bukalapak.....	53
Tabel 4.8 Frekuensi data setelah proses stemming Blibli.com.....	54
Tabel 4.9 Count Vectorized untuk Komentar	55
Tabel 4.10 Frekuensi kata Negatif Blibli.com.	58
Tabel 4.11 Frekuensi kata Negatif Bukalapak.	59
Tabel 4.12 Frekuensi kata Negatif Lazada.....	60
Tabel 4.13 NBC Lazada.....	61
Tabel 4.14 Probabilitas Klasifikasi NBC pada Lazada.....	63
Tabel 4.15 NBC Bukalapak.	64
Tabel 4.16 Probabilitas Klasifikasi NBC pada Bukalapak.	65
Tabel 4.17 NBC Blibli.com	66

Tabel 4.18 Probabilitas Klasifikasi NBC pada Blibli.com.	68
Tabel 4.19 Hasil Perhitungan NBC Pada situs belanja.	69
Tabel 4.20 Ketepatan Klasifikasi metode NBC dengan Menggunakan 10-Fold Cross Validation pada Lazada.	69
Tabel 4.21 Ketepatan Klasifikasi metode NBC dengan Menggunakan 10-Fold Cross Validation pada Bukalapak.	70
Tabel 4.22 Ketepatan Klasifikasi metode NBC dengan Menggunakan 10-Fold Cross Validation pada Blibli.com.	71
Tabel 4.23 Rating secara umum komentar priceprice.	72
Tabel 4.24 Rating secara umum berdasarkan komentar positif.	72
Tabel 4.25 Rating secara umum berdasarkan paling populer.	73
Tabel 4.26 Rating secara umum berdasarkan paling aman.	73
Tabel 4.27 Rating secara umum berdasarkan paling memuaskan.	74
Tabel 4.28 Pemeringkatan Keseluruhan.	75

BAB I

PENDAHULUAN

1.1 Latar Belakang

Menurut lembaga riset pasar *e-Marketer*, populasi *netter* tanah air mencapai 83,7 juta orang pada 2014 (*e-marketer*, 2014). Angka yang berlaku untuk setiap orang yang mengakses internet setidaknya satu kali setiap bulan itu mendudukkan Indonesia di peringkat ke-6 terbesar di dunia dalam hal jumlah pengguna internet. Pada 2017, *eMarketer* memperkirakan *netter* Indonesia bakal mencapai 112 juta orang, mengalahkan Jepang di peringkat ke-5 yang pertumbuhan jumlah pengguna internetnya lebih lamban. Secara keseluruhan, jumlah pengguna internet di seluruh dunia diproyeksikan bakal mencapai 3 miliar orang pada 2015. Tiga tahun setelahnya, pada 2018, diperkirakan sebanyak 3,6 miliar manusia di bumi bakal mengakses internet setidaknya sekali tiap satu bulan.

Top 25 Countries, Ranked by Internet Users, 2013-2018						
millions						
	2013	2014	2015	2016	2017	2018
1. China*	620.7	643.6	669.8	700.1	736.2	777.0
2. US**	246.0	252.9	259.3	264.9	269.7	274.1
3. India	167.2	215.6	252.3	283.8	313.8	346.3
4. Brazil	99.2	107.7	113.7	119.8	123.3	125.9
5. Japan	100.0	102.1	103.6	104.5	105.0	105.4
6. Indonesia	72.8	83.7	93.4	102.8	112.6	123.0
7. Russia	77.5	82.9	87.3	91.4	94.3	96.6
8. Germany	59.5	61.6	62.2	62.5	62.7	62.7
9. Mexico	53.1	59.4	65.1	70.7	75.7	80.4
10. Nigeria	51.8	57.7	63.2	69.1	76.2	84.3
11. UK**	48.8	50.1	51.3	52.4	53.4	54.3
12. France	48.8	49.7	50.5	51.2	51.9	52.5
13. Philippines	42.3	48.0	53.7	59.1	64.5	69.3
14. Turkey	36.6	41.0	44.7	47.7	50.7	53.5
15. Vietnam	36.6	40.5	44.4	48.2	52.1	55.8
16. South Korea	40.1	40.4	40.6	40.7	40.9	41.0
17. Egypt	34.1	36.0	38.3	40.9	43.9	47.4
18. Italy	34.5	35.8	36.2	37.2	37.5	37.7
19. Spain	30.5	31.6	32.3	33.0	33.5	33.9
20. Canada	27.7	28.3	28.8	29.4	29.9	30.4
21. Argentina	25.0	27.1	29.0	29.8	30.5	31.1
22. Colombia	24.2	26.5	28.6	29.4	30.5	31.3
23. Thailand	22.7	24.3	26.0	27.6	29.1	30.6
24. Poland	22.6	22.9	23.3	23.7	24.0	24.3
25. South Africa	20.1	22.7	25.0	27.2	29.2	30.9
Worldwide***	2,692.9	2,892.7	3,072.6	3,246.3	3,419.9	3,600.2

Note: Individuals of any age who use the internet from any location via any device at least once per month; *excludes Hong Kong; **forecast from Aug 2014; ***includes countries not listed
Source: eMarketer, Nov 2014

Gambar 1.1 Pengguna Internet di dunia selama kurun waktu 6 tahun terakhir.

Sumber : comScore MMX Multi-platform, Juni 2017, Indonesia

Definisi *marketplace* berdasarkan ilmu *One Data* adalah fasilitator pembelian online yang tidak memiliki inventarisasi barangnya sendiri. Sedangkan, *e-commerce* merupakan istilah yang digunakan untuk menyebut suatu kegiatan perdagangan melalui sistem elektronik seperti internet ataupun televisi. Gambar 1.2 Situs online di Indonesia dengan 10 peringkat teratas.

Top Performing Online Consumer Goods Retailers in Indonesia



		Total Digital Population (000)	Mobile (000)	Desktop (000)	Total Minutes (MM)	Total Views (MM)	Avg. Minutes per View
1	Lazada.co.id	21,235	15,864	8,107	526	552	1
2	Blibli.com	15,556	13,837	2,651	635	422	1.5
3	Tokopedia.com	14,401	13,006	2,217	1,548	326	4.7
4	Elevenia.co.id	12,872	9,535	5,130	438	285	1.5
5	MatahariMall.com	12,520	11,516	1,879	410	516	0.8
6	Shopee.co.id	11,301	10,872	763	2,169	136	16
7	Bukalapak.com	10,407	8,971	2,203	459	193	2.4
8	Zalora.co.id	9,052	8,636	813	396	493	0.8
9	Qoo10.co.id	7,689	7,641	123	76	91	0.8
10	Blanja.com	5,823	5,673	327	81	88	0.9

Sumber : comScore MMX Multi-platform, Juni 2017, Indonesia


Gambar 1.3 Situs online di Indonesia dengan 10 peringkat teratas.

Dihitung dari jumlah *unique audiens* pada akhir triwulan dua, didapatkan data dari 10 peringkat teratas yang terdiri atas 4 *e-commerce* dan 6 *marketplace*. Lazada memimpin seluruh *e-commerce* dengan 21,2 juta *unique audiens*. Urutan ke-10 peringkat teratas tersebut antara lain Lazada, Blibli, Tokopedia, Elevenia, Matahari Mall, Shopee, Bukalapak, Zalora, Qoo10, dan Blanja. Penulis memutuskan untuk memilih Lazada, Bukalapak, dan Blibli untuk dijadikan objek penelitian dikarenakan banyak terdapat ulasan mengenai tiga situs tersebut pada platform website yang dituju.

Pada kasus ini penulis hanya menekankan pada tiga situs belanja online yakni Lazada, Bukalapak, dan Blibli. Berikut penulis tampilkan kelebihan dan kekurangan dari ketiga situs online yang akan dijadikan objek penelitian:

Tabel 1.1 Kelebihan dan kekurangan pada situs belanja online yang diteliti (Dewantoro,2016).

Situs Online	Kelebihan	Kekurangan
	<ul style="list-style-type: none"> • Beraneka Ragam Produk Tersedia. • Banyak Diskon Menarik Yang Di Tawarkan. • Website yang user friendly • Bebas Ongkos Kirim • COD (Cash On Delivery) • Pengiriman Barang Cepat 	<ul style="list-style-type: none"> • Konfirmasi Pembayaran Cukup Lama. • Fasilitas COD masih belum bisa dibanyak kota.
	<ul style="list-style-type: none"> • Di bukalapak kita bisa menemukan barang-barang bekas. Bukalapak mengizinkan third-party seller/retailer (penjual pihak ketiga) untuk menjual barang bekas, jadi jangan heran jika kita mencari barang yang tidak ada di situs online lain namun tersedia di bukalapak, terutama jika kita mencari barang-barang jadul yang sudah tidak diproduksi. • Fitur terbaiknya, bukalapak menyediakan fitur mobile untuk browser handphone lawas. Jadi ketika seseorang menggunakan browser di handphone lawas, katakanlah itu nokia dengan sistem java, maka tampilan bukalapak akan menjadi tampilan sederhana, sehingga loadingnya akan terasa cepat. 	<ul style="list-style-type: none"> • Respon customer service yang lama, jika situs lain menyediakan nomor serta alamat lengkap kantor mereka untuk handle setiap aduan jika ada masalah dalam transaksi, maka kita akan kesulitan untuk menghubungi pihak bukalapak. • Begitu juga ketika kita menghubungi lewat email, aduan yang saya kirim (menurut dari catatan saya) biasanya akan mendapatkan respon keesokan harinya.

Situs Online	Kelebihan	Kekurangan
	<ul style="list-style-type: none"> • Ada semacam petunjuk cukup rinci mengenai pemakaian aplikasi. Selain itu, aplikasi ini juga bisa digunakan untuk mengisi pulsa dan token PLN secara langsung. • Tampilan antarmuka aplikasi Blibli pun bisa dibilang simpel. • Informasi produk sangat mudah diakses dan tidak perlu repot-repot <i>login</i> terlebih dahulu ketika ingin berbelanja. 	<ul style="list-style-type: none"> • Sistem konfirmasi pembayaran transfer bank yang memakan waktu 1 hari sejak mengirimkan uang jika order di hari kerja dan 2-3 hari jika membelinya di hari sabtu dan minggu. • Tidak ada cara konfirmasi pembayaran manual, tampaknya harus komplain ke customer care Blibli untuk hal ini.

Komentar pengguna akan digunakan oleh penulis sebagai sumber data yang akan diolah untuk dijadikan wawasan dalam penelitian. Komentar pengguna didapat dari *web id.priceprice.com* yang memuat komentar positif maupun negatif. *Priceprice* adalah platform *review e-commerce* yang unggul dan kredibel. Karena itulah penulis menggunakan *platform* terbuka sehingga *review* tersebut sesuai dengan kondisi asli dari produk atau jasa tersebut.

Adapun penelitian-penelitian terdahulu yang terkait dengan *sentiment analysis*, antara lain adalah penelitian (Gamallo, Pablo, & FernandezLanza, 2013) tentang *strategy naïve bayes* untuk *sentiment analysis* untuk *tweet* berbahasa Spanyol. Penelitian (Vinodhini and Chandrasekaran, 2012)

mengembangkan sistem yang dapat mengidentifikasi dan mengklasifikasikan sentimen masyarakat untuk memprediksi produk yang menarik dalam pemasaran. Penelitian (Murnawan, Ardiles Sinaga, 2017) tentang Pemanfaatan Analisis Sentimen Untuk Peningkatan Popularitas Tujuan Wisata.

Pada penelitian ini, *sentiment analysis* akan digunakan untuk menentukan peringkat popularitas tujuan pembelian berdasarkan data komentar atau opini yang terdapat di forum diskusi yang spesifik membicarakan mengenai situs belanja *online*. Dengan harapan dapat mencapai tujuan penelitian dalam penentuan popularitas situs belanja online di Indonesia. Manfaat yang dapat diambil pada penelitian ini, yakni pengguna aplikasi belanja online dapat lebih bijak dalam menentukan situs mana yang aman dalam proses transaksinya. Selain itu, pihak pemilik situs juga dapat meningkatkan kinerja layanannya dengan adanya analisis ini.

Pada penelitian ini juga, peneliti memutuskan untuk menggunakan strategi klasifikasi berdasarkan pada algoritma *Naive Bayes* (NB) karena merupakan suatu metode yang sederhana dan merupakan dan intuitif yang kinerjanya mirip dengan pendekatan lain. Selain itu, berdasarkan penelitian-penelitian terdahulu, NB menggabungkan efisiensi (waktu kinerja yang optimal) dengan cukup akurat.

1.2 Rumusan Masalah

Dari uraian latar belakang diatas dapat dirumuskan bahwa masalah yang ada yaitu bagaimana tingkat popularitas situs belanja online di Indonesia diantara Lazada, Bukalapak dan Blibli?

1.3 Batasan Masalah

Berdasarkan dari rumusan masalah diatas, maka penulis menentukan batasan masalah sebagai berikut :

1. Analisis ini hanya membahas tiga situs online di Indonesia, yakni Lazada, Bukalapak dan Blibli.
2. Komentar pengguna didapat dari *web id.priceprice.com* yang memuat komentar positif maupun negatif.

1.4 Tujuan

Adapun tujuan dari penelitian ini adalah untuk menentukan situs belanja *online* terpopuler di Indonesia diantara Lazada, Bukalapak dan Blibli.

1.5 Manfaat

Berdasarkan analisis pemilihan situs belanja *online* terbaik di Indonesia ini, maka pengguna aplikasi penjualan online dapat memilih dengan bijak manakah situs belanja yang aman dalam proses transaksinya. Kemudian pada pihak pemilik situs belanja juga dapat meningkatkan kinerja layanan mereka dengan adanya analisis ini.

BAB II

KAJIAN PUSTAKA

2.1 E-Commerce

E-commerce merupakan suatu istilah yang sering digunakan atau didengar saat ini yang berhubungan dengan internet, dimana tidak seorangpun yang mengetahui jelas pengertian dari *e-commerce* tersebut. Berikut akan dipaparkan pengertian *e-commerce* menurut para ahli :

- a. Perdagangan elektronik atau yang disebut juga *e-commerce*, adalah penggunaan jaringan komunikasi dan komputer untuk melaksanakan proses bisnis. Pandangan populer dari *e-commerce* adalah penggunaan internet dan komputer dengan browser *Web* untuk membeli dan menjual produk. McLeod Pearson (2008 : 59).
- b. Menurut Jony Wong (2010 : 33) pengertian dari *electronic commerce* adalah pembelian, penjualan dan pemasaran barang serta jasa melalui sistem elektronik. Seperti radio, televisi dan jaringan computer atau internet.

2.2 Analisis

Analisis adalah sebuah teknik pemecahan masalah yang menguraikan sebuah sistem menjadi bagian – bagian komponen untuk tujuan mempelajari seberapa baik komponen tersebut bekerja dan berinteraksi untuk mencapai tujuannya (Whitten J. L., 2007). Menurut Wiradi, analisis adalah aktivitas yang memuat sejumlah kegiatan seperti mengurai, membedakan, memilah sesuatu untuk digolongkan dan dikelompokkan kembali menurut kriteria tertentu kemudian dicari kaitannya dan ditafsir maknanya (Makinudin, 2006).

2.3 Sentiment Analisis dan Text Mining

Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, kalimat, atau fitur/tingkat aspek dan menentukan apakah pendapat yang dikemukakan dalam dokumen, kalimat atau fitur entitas/aspek bersifat positif, negatif atau netral. Lebih lanjut *sentiment analysis* dapat menyatakan emosional sedih, gembira, atau marah (Liu, 2012).

Ekspresi atau sentiment mengacu pada fokus topik tertentu, pernyataan pada satu topik mungkin akan berbeda makna dengan pernyataan yang sama pada subject yang berbeda. Bahkan pada produk tertentu, kata-kata yang sama dapat menggambarkan makna kebalikan, contoh adalah hal yang buruk untuk waktu start-up pada kamera digital jika dinyatakan “lama”, namun jika “lama” dinyatakan pada usia baterai maka akan menjadi hal positif.

Hal pertama dalam pemrosesan dokumen adalah memecah kumpulan karakter ke dalam kata atau token, sering disebut sebagai tokenisasi. Tokenisasi adalah hal yang kompleks untuk program komputer karena beberapa karakter dapat ditemukan sebagai token delimiters. Delimiter adalah karakter spasi, tab dan baris baru (newline), sedangkan karakter () < > ! ? “ kadang dijadikan delimiter namun, kadang juga bukan, tergantung pada lingkungannya (Triawati, 2009).

Text mining merupakan penerapan konsep dan teknik data mining untuk mencari pola dalam teks, yaitu proses penganalisisan teks guna menyarikan informasi yang bermanfaat untuk tujuan tertentu. Berdasarkan ketidakaturan struktur data teks, maka proses *text mining* memerlukan beberapa tahap awal yang pada intinya adalah mempersiapkan agar teks dapat diubah menjadi lebih terstruktur.

Tujuan dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Jadi, sumber data yang digunakan pada *text mining* adalah kumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur. Adapun tugas khusus dari *text mining* antara lain yaitu pengkategorisasian teks (*text categorization*) dan pengelompokan teks (*text clustering*).

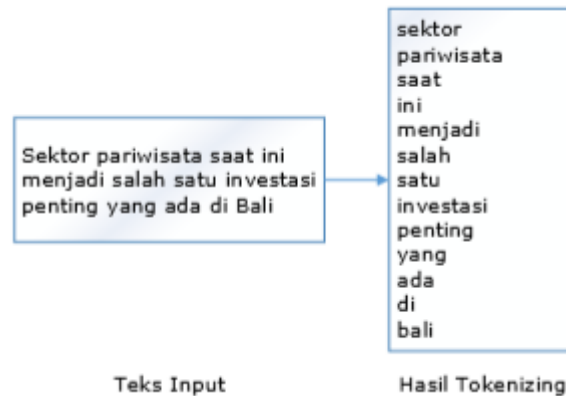
Proses *text preprocessing* merupakan implementasi dari *text mining* untuk memproses suatu teks yang tidak terstruktur menjadi lebih terstruktur atau dengan kata lain mempersiapkan agar teks dapat diubah menjadi lebih terstruktur. Ada empat tahapan *text preprocessing* yang meliputi *case folding*, *tokenizing*, *filtering*, dan *stemming*.

1) *Case folding*

Tahapan *case folding* adalah tahapan untuk mengkonversi keseluruhan teks dalam dokumen menjadi seluruhnya huruf kecil (*lowercase*). Sebagai contoh, teks “PARIWISATA”, “Pariwisata”, “PariWisata”, atau “pariwisata”, tetap diberikan hasil *retrieval* yang sama yakni “pariwisata”. Pada tahapan *case folding* ini hanya huruf ‘a’ sampai dengan ‘z’ yang diterima, sedangkan karakter selain huruf dihilangkan dan dianggap delimiter.

2) *Tokenizing*

Tahapan *tokenizing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Contoh dari tahap ini dapat dilihat pada Gambar 2.1 dibawah ini.

Gambar 2.1 Tahap *Tokenizing*

sekumpulan karakter dalam suatu teks ke dalam satuan kata, bagaimana membedakan karakter-karakter tertentu yang dapat diperlakukan sebagai pemisah kata atau bukan. Sebagai contoh karakter *whitespace*, seperti enter, tabulasi, spasi dianggap sebagai pemisah kata. Namun untuk karakter petik tunggal (‘), titik (.), semikolon (;), titik dua (:) atau lainnya, dapat memiliki peran yang cukup banyak sebagai pemisah kata.

Dalam memperlakukan karakter-karakter dalam teks sangat tergantung pada konteks aplikasi yang dikembangkan. Pekerjaan tokenisasi ini akan semakin sulit jika juga harus memperhatikan struktur bahasa (grammatikal).

3) *Filtering*

Tahapan *filtering* adalah tahap mengambil kata-kata penting dari hasil token. Bisa menggunakan algoritma *stoplist* (membuang kata kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist/stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*.



Gambar 2.2 Tahap Filtering

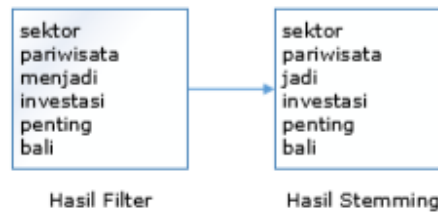
Kata-kata seperti “dari”, “yang”, “di”, dan “ke” adalah beberapa contoh kata-kata yang berfrekuensi tinggi dan dapat ditemukan hampir dalam setiap dokumen (disebut sebagai *stopword*). Penghilangan *stopword* ini dapat mengurangi ukuran index dan waktu pemrosesan. Namun terkadang *stopping* tidak selalu meningkatkan nilai *retrieval*. Pembangunan daftar *stopword* (disebut *stoplist*) yang kurang hati-hati dapat memperburuk kinerja sistem *Information Retrieval (IR)*. Belum ada suatu kesimpulan pasti bahwa penggunaan *stopping* akan selalu meningkatkan nilai *retrieval*, karena pada beberapa penelitian, hasil yang didapatkan cenderung bervariasi.

4) *Stemming*

Pembuatan indeks dilakukan karena suatu dokumen tidak dapat dikenali langsung oleh suatu sistem temu kembali informasi atau *Information Retrieval System (IRS)*. Oleh karena itu, dokumen tersebut terlebih dahulu perlu dipetakan ke dalam suatu representasi dengan menggunakan teks yang berada di dalamnya.

Teknik *stemming* diperlukan selain untuk memperkecil jumlah indeks yang berbeda dari suatu dokumen, juga untuk melakukan pengelompokan kata-kata lain

yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk atau form yang berbeda karena mendapatkan imbuhan yang berbeda.



Gambar 2.3 Tahap Stemming

Sebagai contoh kata bersama, kebersamaan, menyamai, akan distem ke *root word*-nya yaitu “sama”. Namun, seperti halnya stopping, kinerja *stemming* juga bervariasi dan sering tergantung pada domain bahasa yang digunakan. Proses *stemming* pada teks berbahasa Indonesia berbeda dengan *stemming* pada teks berbahasa Inggris. Pada teks berbahasa Inggris, proses yang diperlukan hanya proses menghilangkan *sufiks*. Sedangkan pada teks berbahasa Indonesia semua kata imbuhan baik itu *sufiks* dan *prefiks* juga dihilangkan. Pada kasus ini menggunakan kamus dari *Liu's word list*.

2.4 Naive Bayes Classifier

Naïve Bayes Classifier merupakan teknik pembelajaran algoritma data mining yang memanfaatkan metode probabilitas dan statistik. *Naïve Bayes Classifier* dalam melakukan klasifikasi terdapat dua proses penting yaitu *learning (training)* dan *testing*.

Naive Bayes Classifier bekerja sangat baik dibanding dengan model *classifier* lainnya. Hal ini dibuktikan pada jurnal Xhemali, Daniela, Chris J. Hinde, and Roger G. Stone. “Naive Bayes vs. decision trees vs. neural networks in the classification of training web pages.” (2009), mengatakan bahwa “*Naïve Bayes*

Classifier memiliki tingkat akurasi yg lebih baik dibanding model *classifier* lainnya”.

Metode *Naive Bayes* digunakan karena lebih cocok dalam pengoperasian analisis sentimen pada penelitian ini dibandingkan dengan metode lainnya seperti *K-Nearest Neighbour* atau *learning machine*. *K-nearest neighbors* atau knn adalah algoritma yang berfungsi untuk melakukan klasifikasi suatu data berdasarkan data pembelajaran (*train data sets*), yang diambil dari k tetangga terdekatnya (*nearest neighbors*). Dengan menggunakan data, pembelajaran mesin memungkinkan komputer menemukan wawasan tersembunyi tanpa diprogram secara eksplisit saat mencarinya. Dengan adanya email baru, algoritma tersebut kemudian akan menghasilkan prediksi apakah email baru itu *spam* atau tidak.

Proses klasifikasi *Naïve Bayes Classifier* terhadap dokumen yaitu dengan mempresentasikan setiap dokumen dengan atribut “ $X_1, X_2, X_3, \dots, X_n$ ” yang mempunyai makna bahwa X_1 untuk kata pertama, X_2 adalah kata kedua, dan seterusnya. Untuk himpunan kategori komentar dipresentasikan dengan V . Saat melakukan proses klasifikasi dokumen, *Naïve Bayes Classifier* akan mencari nilai probabilitas tertinggi dari [4]:

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} \frac{P(x_1, x_2, x_3, \dots, x_n | V_j) P(V_j)}{P(x_1, x_2, x_3, \dots, x_n)}$$

Keterangan:

V_{MAP} : semua kategori yang diujikan

V_j : Kategori komentar, dengan:

J_1 : Sentimen positif

J_2 : Sentimen Negatif

$P(X_i/V_j)$: probabilitas X_i pada kategori V_j

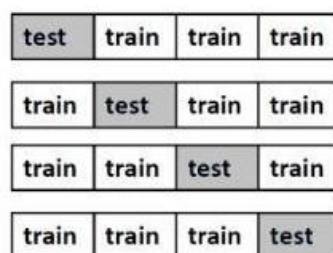
$P(V_j)$: probabilitas dari V_j

Proses Klasifikasi *Naïve Bayes*

Tujuan dari proses klasifikasi adalah untuk menentukan sebuah kalimat apakah termasuk sebagai anggota kelas opini positif, negatif atau sebagai anggota kelas opini netral yang ditentukan berdasarkan nilai perhitungan probabilitas Bayes yang lebih besar. Jika hasil probabilitas *Bayes* kalimat tersebut untuk kelas opini positif lebih besar maka kalimat tersebut masuk kategori opini positif demikian juga sebaliknya.

10-Fold Cross-Validation

Model yang didapatkan dari *naïve bayes* kemudian dilakukan pengujian menggunakan *10-fold cross validation*. *Cross-validation* adalah bentuk sederhana dari teknik statistik. Jumlah fold standar untuk memprediksi tingkat error dari data adalah dengan menggunakan *10-fold cross validation* (Witten, et al, 2011: 153). Data yang digunakan dibagi secara acak ke dalam k subset yaitu D_1, D_2, \dots, D_k dengan ukuran yang sama. Dataset akan dibagi menjadi data *training* dan data *testing*. Proses training dan testing dilakukan sebanyak k kali secara berulang-ulang. Pada iterasi ke-i, partisi D_i disajikan sebagai data testing dan partisi sisanya digunakan secara bersamaan dan berurutan sebagai data *training*.



Gambar 2.4 Ilustrasi 4-fold cross-validation

Berdasarkan Gambar 2.4. ditunjukkan bahwa nilai *fold* yang digunakan adalah *4-fold cross validation*. Berikut diberikan langkah-langkah pengujian data dengan *4-fold cross validation*. Untuk menguji keakuratan data aplikasi yang digunakan pada metode ini memanfaatkan aplikasi Rapidminer.

2.5 Lazada



Gambar 2.5 Logo Lazada

Bisnis online di era serba cepat dan mudah sangatlah menguntungkan. Bisnis *online* yang menjual barang di internet salah satunya adalah situs Lazada.com. Situs ini bergerak dalam bidang IT sebagai penyedia barang seperti elektronik, buku, mainan anak dan perlengkapan bayi, alat kesehatan dan produk kecantikan, peralatan rumah tangga, dan perlengkapan traveling dan olah raga. Lazada Indonesia didirikan pada tahun 2012 dan merupakan salah satu cabang dari jaringan *retail online* lazada di Asia Tenggara. Grup lazada International di Asia Tenggara. Pada Januari 2012, jumlah karyawan lazada hanya 4 orang, dan pada bulan Agustus 2012 karyawan yang dipekerjakan naik menjadi 200 karyawan dan terus bertambah setiap bulannya.

Tom Damek selaku *Managing Director* di Indonesia enggan menyebutkan jumlah omset yang mereka dapatkan selama ini namun menurut tafsiran pengamat, omset lazada dalam sebulan mencapai milyaran rupiah. Lazada mempunyai *warehouse* atau gudang seluas 2.500 meter persegi dan 5.000 meter persegi di

Jakarta, hal ini merupakan upaya untuk mempertahankan pertumbuhan sebesar 20% setiap bulan. Hingga Desember 2012, lazada memiliki 13 kategori produk, dan lebih dari 3.000 subkategori. Total, perusahaan menjual 25.000 jenis produk. Pada tahun lalu, Lazada mendapatkan beberapa kali suntikan dana dari beberapa perusahaan seperti JP Mogan dan Kinnevik. Selain Indonesia, Lazada juga beroperasi di empat negara di Asia Tenggara, yakni Malaysia, Thailand, Vietnam, dan Filipina.

Sedangkan untuk kekurangan dari *E-commerce* Lazada Indonesia adalah :

1. Tidak menutup kemungkinan dari *costumer* yang berkunjung ke *web* ini dan meragukan integritas dari situs *E-commerce* ini yang dikarenakan nomor *customer* yang dicantumkan di situs merupakan nomor *provider* untuk telepon genggam yang dianggap sama kebanyakan situs / penjual produk bohongan (penipuan) yang juga tersebar di internet.
2. Respon layanan konfirmasi pembayaran yang masih sedikit terlambat.
3. Beberapa produk yang available stock di website ternyata *out of stock* di gudang.
4. Layanan *M-Banking* dan pembayaran kartu kredit yang belum memiliki integritas tinggi.

2.6 Bukalapak



Gambar 2.6 Logo Bukalapak

Bukalapak merupakan salah satu pasar daring (*online marketplace*) terkemuka di Indonesia yang dimiliki dan dijalankan oleh PT. Bukalapak. Seperti halnya situs layanan jual-beli online dengan model *bisnis consumer-to-consumer* (C2C), Bukalapak menyediakan sarana penjualan dari konsumen ke konsumen di mana pun. Siapa pun bisa membuka toko daring untuk kemudian melayani calon pembeli dari seluruh Indonesia baik satuan ataupun dalam jumlah banyak. Pengguna perorangan ataupun perusahaan dapat membeli dan menjual produk, baik baru maupun bekas, seperti sepeda, ponsel, perlengkapan bayi, gadget, aksesoris gadget, komputer, tablet, perlengkapan rumah tangga, busana, elektronik, dan lain-lain. (Silviosilviando, 2016)

Bukalapak didirikan oleh Achmad Zaky pada awal tahun 2010 sebagai divisi agensi *digital* bernama Suitmedia yang berbasis di Jakarta. Namun, Bukalapak baru berstatus sebagai sebuah Perseroan Terbatas (PT) pada September 2011 dan dikelola oleh manajemen yang dipimpin oleh Achmad Zaky sebagai *CEO* (*Chief Executive Office*) dan Nugroho Herucahyono sebagai *CTO* (*Chief Technology Officer*).

Setelah berdiri kurang lebih setahun, Bukalapak mendapat penambahan modal dari *Batavia Incubator* (perusahaan gabungan dari Rebright Partners yang dipimpin oleh Takeshi Ebihara, Japanese Incubator dan Corfina Group). Di tahun 2012,

Pada bulan Maret 2014, Bukalapak mengumumkan investasi oleh Aucfan, IREP, 500 Startups, dan GREE Ventures. Tidak berselang lama dari pemberitaan tersebut, di tanggal 18 Maret 2014 Bukalapak pun meluncurkan aplikasi selular untuk Android. Sejak pertama kali diluncurkan sampai dengan Oktober 2016, aplikasi tersebut telah didownload oleh lebih dari 5 juta kali dan masuk dalam 100 Aplikasi Gratis terpopuler di Google Play.

Bukalapak memiliki program untuk memfasilitasi para UKM yang ada di Indonesia untuk melakukan transaksi jual beli secara *online*. Hal ini dikarenakan transaksi melalui online dapat mempermudah UKM dalam menjual produk-produk yang mereka miliki tanpa harus memiliki toko offline. Untuk yang telah memiliki toko offline, Bukalapak mengharapkan dengan adanya situs tersebut dapat membantu meningkatkan penjualan toko *offline* tersebut.

2.7 Blibli.com



Gambar 2.7 Logo Blibli.com

Blibli.com diluncurkan pada 25 Juli 2011 oleh PT. Global Digital Niaga (GDN) yang merupakan anak perusahaan PT. Global Digital Prima (GDP) milik Martin

Hartono –putra owner Djarum Group, Budi Hartono. PT. Global Digital Prima (GDP) adalah bagian dari perusahaan produsen rokok terbesar di Indonesia Djarum Group yang juga menjadi pemegang saham beberapa bisnis online terkemuka seperti situs *e-commerce* macam lintasberita.com, krazymarket.com, dailysocial.net, dan perusahaan inkubator bisnis online bernama Merah Putih Inc. (Ervinzein, 2016)

Sejak lulus SMA pada 1992, Kusumo Martanto melanjutkan pendidikan di AS dan bekerja di beberapa perusahaan di AS seperti Intel Corporation dan i2 technologies dengan jabatan terakhir senior manajer produk. Baru pada tahun 2009 dia pulang ke Indonesia karena pinangan Djarum Group dengan memulai debut karirnya lewat Blibli.com dibawah naungan PT. Global Digital Prima (GDP).

Visi Perusahaan Blibli.com merupakan *social e-commerce* yang dikembangkan dengan visi untuk memberikan pengalaman berbelanja online yang aman, nyaman, dan membebaskan. Misi Perusahaan Blibli.com memiliki misi untuk menghadirkan layanan social commerce yang didasarkan pada nilai-nilai:

1. Menjunjung semangat kebersamaan
2. Transparan
3. Sesuai dengan norma ke-Indonesiaan
4. Komitmen pada pelayanan.

Strategi Perusahaan Strategi Kompetitif Menentukan Target Pasar yang Kuat
Membangun Relasi dan Memperbanyak Variant Produk Strategi Ekspansi.
Pengadaan fitur dalam website untuk menarik konsumen Promosi Cicilan 0%
Semua Produk Blibli *Reward* Pengiriman Gratis.

2.8 Rapid Miner

Rapid Miner adalah aplikasi data mining yang tidak perlu dipertanyakan lagi dan berbasis sistem *open-source*. Tersedia sebagai aplikasi yang berdiri sendiri untuk analisis data dan sebagai mesin data mining untuk integrasi ke dalam produk sendiri. Ribuan aplikasi RapidMiner di lebih dari 40 negara memberikan pengguna mereka keunggulan yang kompetitif. Solusi yang di usung antara lain :Integrasi data, Analitis ETL, Data Analisis, dan Pelaporan dalam satu suite tunggal.

2.9 Penelitian Sebelumnya

Penelitian sebelumnya, penulis melakukan tinjauan terhadap beberapa penelitian sebelumnya sesuai dengan topik tugas akhir yang diambil sebagai referensi pada penelitian ini :

Tabel 2.1 Penelitian sebelumnya

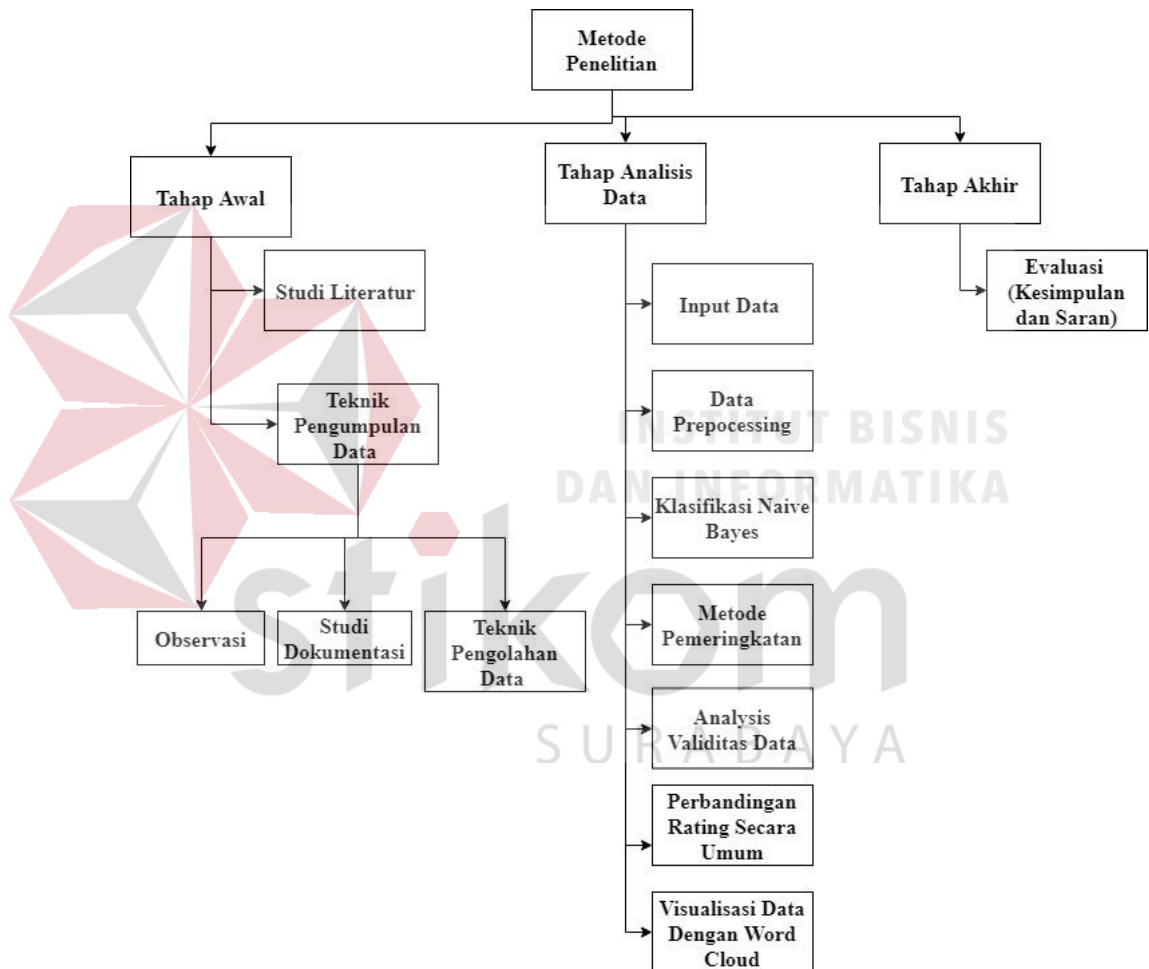
No	Nama Penulis	Judul	Hasil
1	Gamallo, Pablo, & FernandezLanza, 2013	TASS: A Naive-Bayes strategy for sentiment analysis on Spanish tweets	Dalam makalah ini, menyediakan survei dan studi komparatif teknik yang ada untuk penambangan opini termasuk pembelajaran mesin dan pendekatan berbasis leksikon, bersama dengan lintas domain dan metode lintas-bahasa dan beberapa metrik evaluasi. Hasil penelitian menunjukkan bahwa metode pembelajaran mesin, seperti SVM dan <i>naïve Bayes</i> memiliki akurasi tertinggi dan dapat dianggap sebagai metode pembelajaran dasar, sedangkan metode berbasis leksikon sangat efektif dalam beberapa kasus, yang memerlukan sedikit upaya dalam dokumen yang berlabel manusia. Dapat dipelajari efek dari berbagai fitur pada penggolong. Dapat disimpulkan bahwa semakin banyak data yang lebih

No	Nama Penulis	Judul	Hasil
			bersih, hasil yang lebih akurat dapat diperoleh.
2	Rizqiyah, Putri, 2018	Analisa Sentimen Menggunakan Naïve Bayes Untuk Mengetahui Presentase Komentar Pada Aplikasi Go-Jek.	Penelitian ini bertujuan untuk menghitung presentase komentar yang diberikan pengguna pada aplikasi Go-jek, sehingga pengguna dapat mengetahui faktor apa yang mengharuskan pengguna menggunakan aplikasi Go-jek dan juga membantu pihak perusahaan dalam merancang pengembangan aplikasi di masa depan.
3.	Murnawan, Ardiles Sinaga, 2017	Pemanfaatan Analisis Sentimen Untuk Peningkatan Popularitas Tujuan Wisata	Untuk menghasilkan informasi popularitas tujuan wisata, sentiment analysis diimplementasikan secara web based dengan dua halaman utama yang paling penting, yaitu halaman informasi peringkat, serta informasi detail dari tujuan wisata 5. Kemudian dapat dipilih salah satu tempat tujuan wisata jika ingin melihat detail informasinya. Adapun detail informasi yang ditampilkan adalah posisi tujuan wisata yang terletak pada peta (map) serta informasi dari hasil crawling dan sentiment analisis.

BAB III

METODOLOGI PENELITIAN

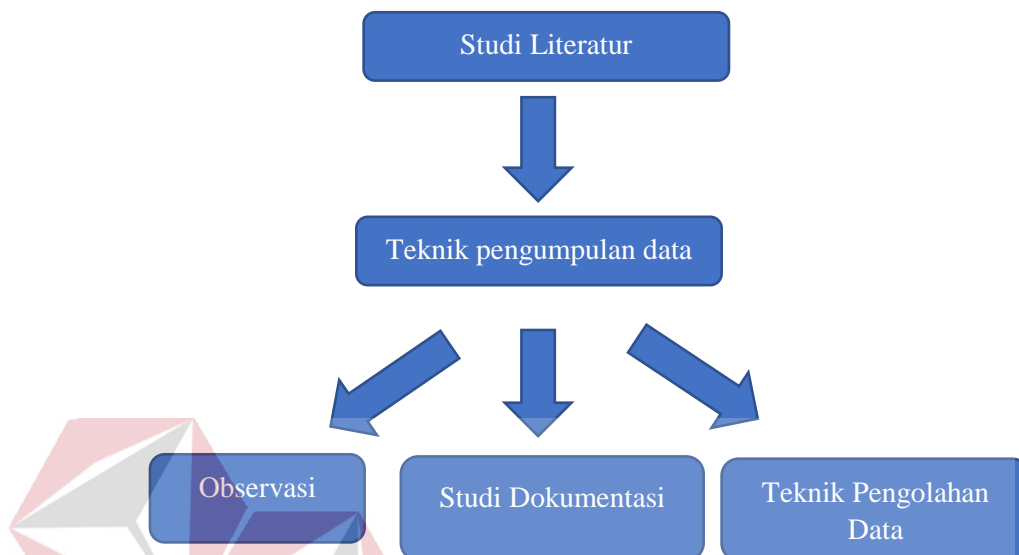
Dalam tahapan penelitian ini akan melalui tiga tahapan yaitu tahap awal penelitian, tahap analisis dan tahap akhir. Dapat dilihat pada gambar 3.1 dibawah ini.



Gambar 3.1 Diagram metode penelitian

3.1 Tahapan Awal Penelitian

Hal yang dilakukan pada penelitian ini:



Gambar 3.2 Bagan Tahapan Awal

Gambar diatas merupakan tahapan penelitian yang penulis lakukan dalam melakukan penelitian ini adapun rincian jelasnya adalah sebagai berikut:

3.1.1.Studi Literatur

Studi literatur ini dilakukan untuk mengetahui model penelitian yang terjadi pada bidang analisis sentimen, studi literatur ini dilakukan untuk mengulas jurnal-jurnal yang telah ada dan digunakan untuk menentukan studi kasus serta metode yang akan digunakan oleh penulis saat melakukan penelitian.

3.1.2 Teknik pengumpulan Data

Pengumpulan data ini dilakukan untuk memperkuat alasan mengapa penelitian ini harus dilakukan, pengumpulan data diambil dari

sumbernya langsung yaitu *website* resmi *id.priceprice.com* serta komentar pengguna aplikasi Lazada, Bukalapak, dan Blibli.com.

Berikut adalah tahapan pengumpulan data yang penulis lakukan dalam menyusun penelitian ini:

A. Observasi

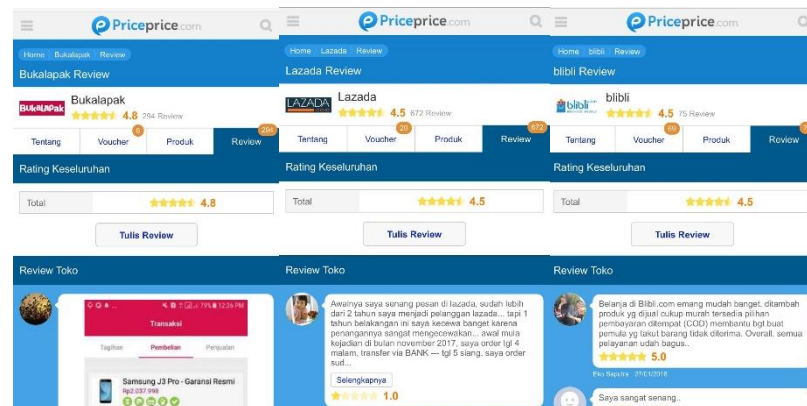
Observasi ini dilakukan untuk mengamati objek yang akan diteliti. Pada penelitian ini dilakukan dengan mengamati komentar pengguna situs belanja *online*, adanya banyak macam komentar yang diberikan oleh pengguna pada aplikasi ini yang membuat penulis melakukan penelitian dalam presentase klasifikasi komentar masyarakat yang ada pada *website id.priceprice.com*.

B. Studi Dokumentasi

Dalam studi dokumentasi ini penulis mengumpulkan data sebagai berikut:

- a. Kurva pengguna dan transaksi aplikasi Lazada, Bukalapak dan Blibli.com.
- b. Klasifikasi ranking yang diberikan pengguna pada aplikasi Lazada, Bukalapak dan Blibli.com.

c. Ranking aplikasi Lazada, Bukalapak dan Blibli.com.



Gambar 3.3 Website priceprice komentar pengguna

C. Teknik Pengolahan Data

Data awal yang dibutuhkan pada penelitian ini adalah *link* dari *marketplace* yang dijadikan sebagai objek. Data *link* yang digunakan pada penelitian ini berbentuk *excel* (.xls). Data *excel* yang didapat akan diolah pada aplikasi *RapidMiner* untuk mengetahui hasil selanjutnya. Berikut dapat dilihat gambar dari data *excel* yang akan diolah pada aplikasi *RapidMiner*. Berikut adalah langkah-langkah dalam teknik pengolahan data :

Data yang diambil adalah data komentar positif maupun negatif dari pengguna *website priceprice.com*, pada tahap ini untuk melakukan pengolahan pada *Rapidminer* yang diambil adalah kode html dari *marketplace* yang dituju. Karena ada tiga objek yang akan diamati pada penelitian ini, maka didapatkan :

1. <http://id.priceprice.com/merchants/lazada-35/reviews/>
2. <http://id.priceprice.com/merchants/Bukalapak-82/reviews/>
3. <http://id.priceprice.com/merchants/blibli-11/reviews/>

Perlu diperhatikan setelah melakukan teknik pengolahan data di atas sebelum melakukan langkah-langkah selanjutnya, komentar yang telah didapatkan dari ketiga *online shop* harus memiliki kata yang baku. Terdapat beberapa kata sehari-hari yang tampil pada penelitian ini, oleh karenanya penulis merubah kata-kata sehari-hari tersebut kedalam kata baku sesuai dengan kamus yang digunakan yakni *Liu's word list*. Pada penelitian ini menggunakan metode perubahan secara manual karena kata yang ditemukan hanya sedikit. Berikut adalah kata yang di temukan dan diubah kedalam kata baku.

Tabel 3.1 Merubah kata menjadi baku.

Kata sehari-hari	Kata baku
Udah	Sudah
Dn	Dan
Drpd	Dari pada
Engga	Tidak
Nggk	Tidak
Gw	Saya
Bgt	Banget
Emang	Memang
Ga	Tidak
Tp	Tapi
Kdg	Kadang
Gak	Tidak

D. Variabel Penelitian.

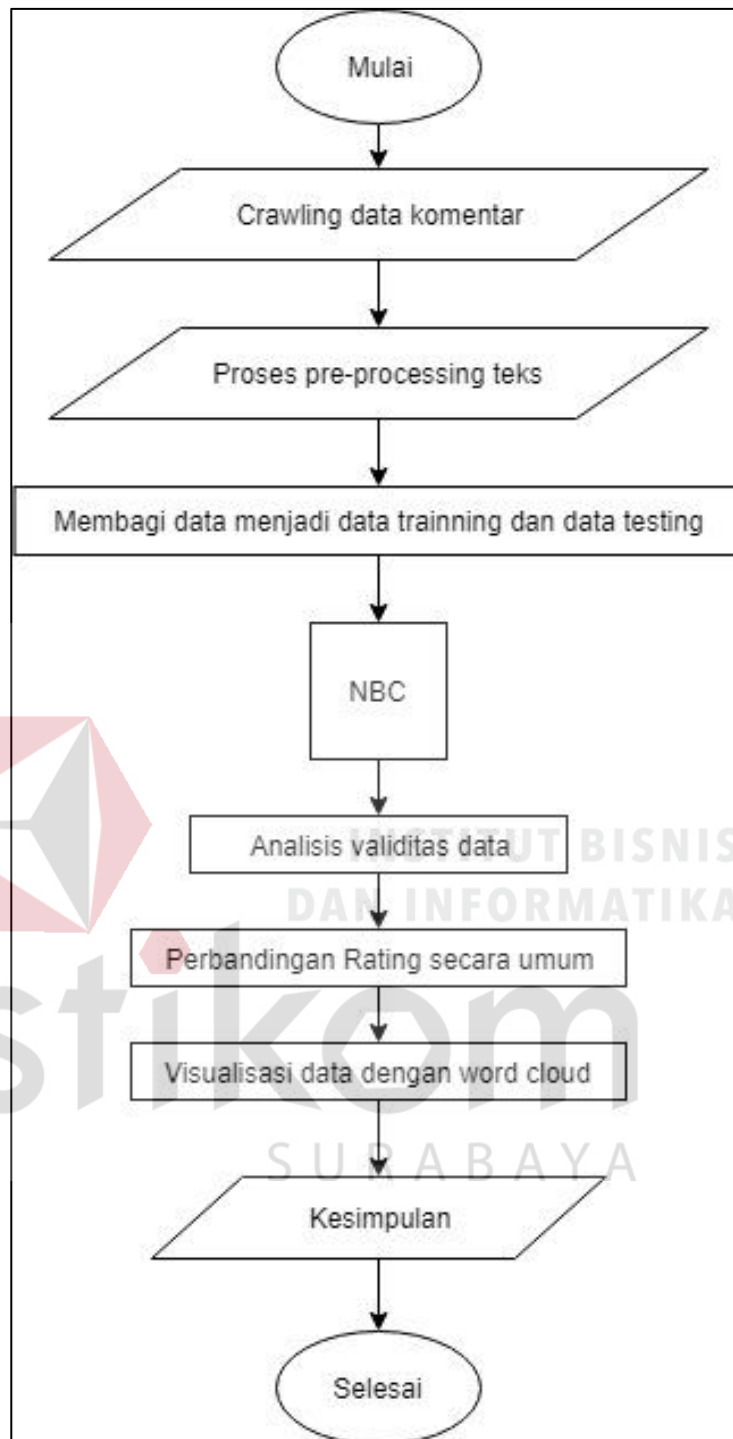
Data yang digunakan dalam penelitian ini dibagi menjadi data *training* sebanyak 90% dan data *testing* sebanyak 10% menggunakan *10-fold cross validation*. Variabel penelitian yang digunakan dalam penelitian ini diberikan dalam Tabel 3.1

Tabel 3.2 Variabel penelitian

Variabel	Keterangan	Skala Data
Y	Sentiment (Positif/Negatif) 0 = Sentiment Positif 1 = Sentimen Negatif	Nominal
X_k	Kata Kunci (Frekuensi Kemunculan Kata ke-k pada komentar ke-n)	Rasio

E. Diagram Alir

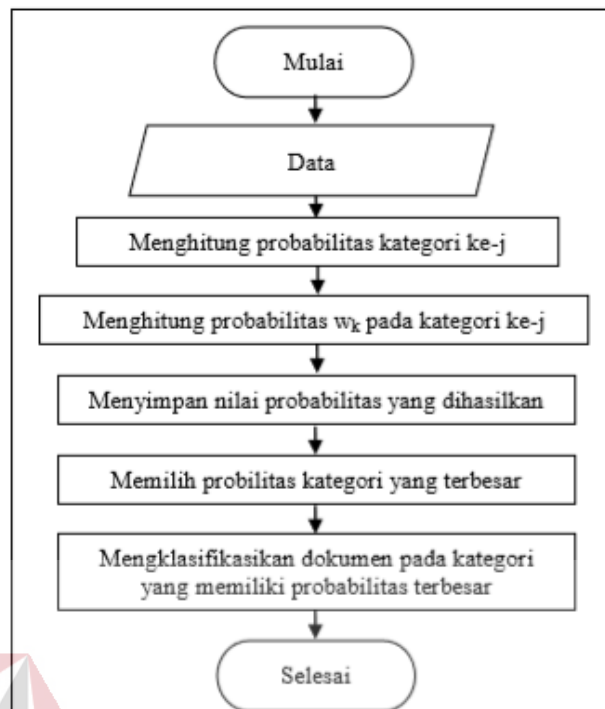
Langkah-langkah analisis ini diberikan dalam bentuk diagram alir pada Gambar 3.4 dan Gambar 3.5



Gambar 3.4 Diagram alir penelitian

Gambar diatas menjelaskan alur dari kegiatan dalam penelitian ini mulai dari *crawling* data komentar pada platform *website priceprice* hingga membuat kesimpulan.

NBC (*Naive Bayes Classification*)



Gambar 3.5 Diagram alir nbc

Gambar 3.5 menjelaskan alur dari klasifikasi *naive bayes* mulai dari data hingga mengklasifikasikan dokumen pada kategori yang memiliki probabilitas terbesar.

3.2 Tahap Analisis Data

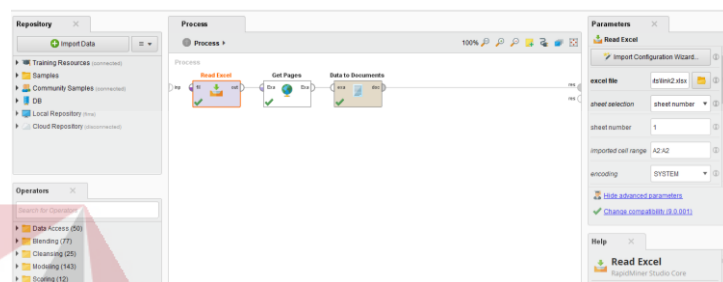
Bagan tahapan analisis data :



Gambar 3.6 Teknik Analisis Data

3.2.1 Input Data

1. Setelah data url didapatkan langkah selanjutnya adalah menyimpan data url tersebut kedalam bentuk *excel*. Pada penelitian ini file terbagi menjadi tiga, yakni lazada, bukalapak dan blibli.
2. Masuk pada aplikasi *rapidminer*. Pilih *read excel* → *Get Pages* → *data to documents*. Hasil dari proses ini adalah sekumpulan kata yang sudah dicrawling dari *website priceprice.com*.



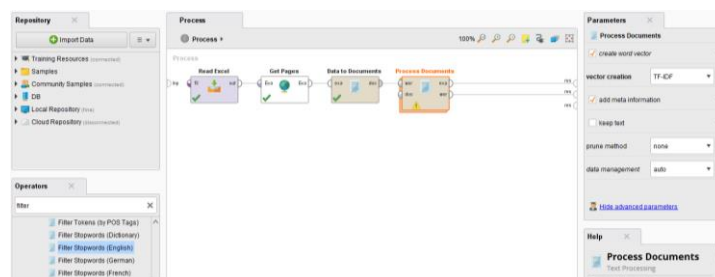
Gambar 3.7 Rapidminer input data.

3. Selanjutnya, sekumpulan data tersebut akan diinputkan pada tahap selanjutnya, yakni pada tahap data *preprocessing*.

3.2.2 Data Preprocessing

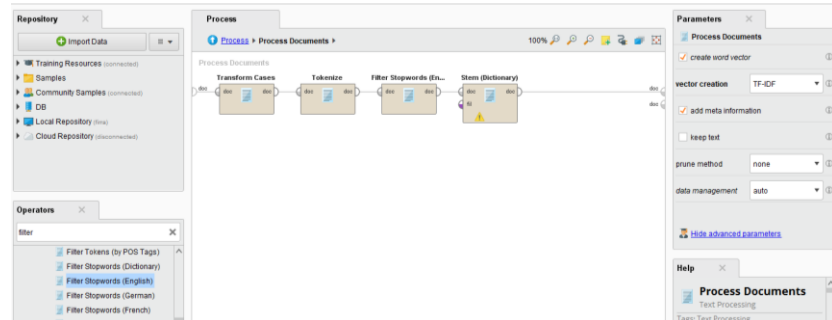
Berikutnya langkah yang harus dilakukan adalah :

1. Lakukan pemrosesan data yang telah diinputkan. Pilih *process documents*.



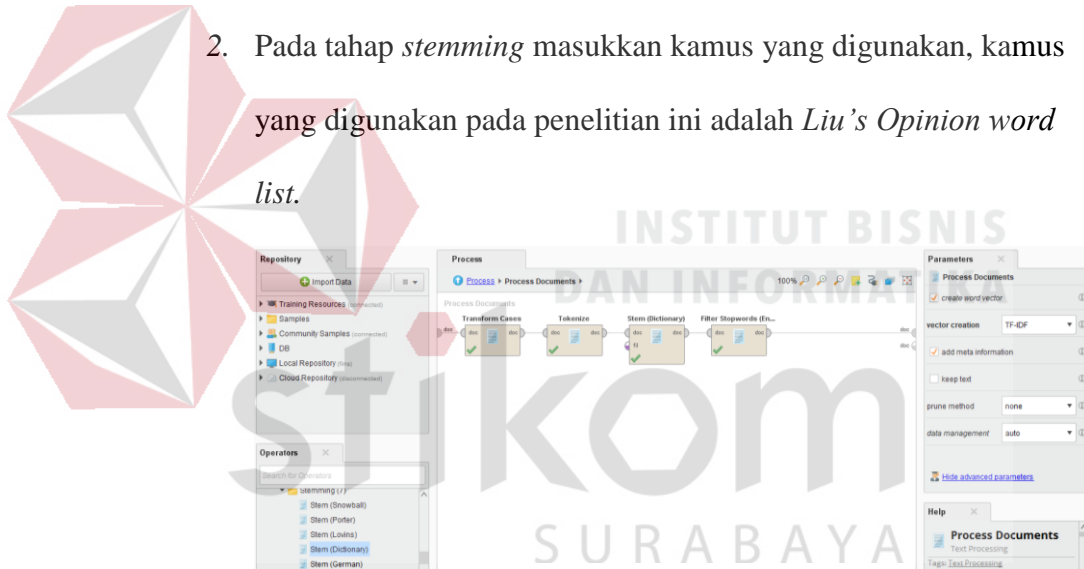
Gambar 3.8 Data preprocessing.

1. Di dalam *process document* masukkan operator yang akan dijalankan. Pada tahap ini tahapan *text preprocessing* yang dilakukan adalah *case folding*, *tokenizing*, *filtering*, dan *stemming*.



Gambar 3.9 Tahapan data preprocessing.

2. Pada tahap *stemming* masukkan kamus yang digunakan, kamus yang digunakan pada penelitian ini adalah *Liu's Opinion word list*.



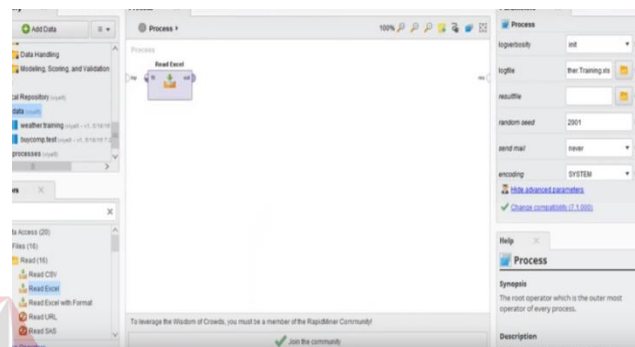
Gambar 3.10 Memasukkan kamus Liu's word list.

3. Hasil dari proses ini adalah suatu file berbentuk *excel* (.xls) dimana akan dibedakan setiap filenya sesuai dengan nama situs belanja dalam penelitian ini, yakni Lazada, bukalapak dan blibli.
4. Masing-masing file akan diklasifikasikan pada proses selanjutnya, yakni proses klasifikasi *naive bayes*. Untuk mengetahui tingkat akurasi data.

3.2.3 Klasifikasi Naive Bayes

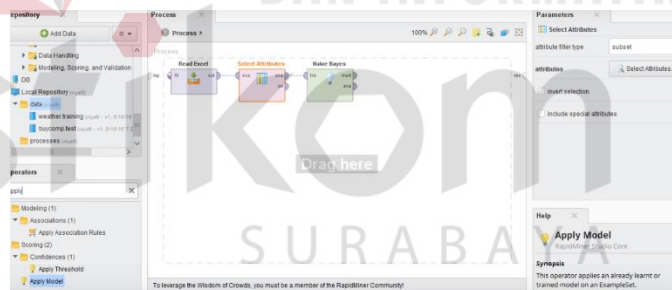
Langkah yang harus dilakukan pada tahap ini adalah :

1. Pilih *read excel* untuk membaca hasil dari stemming yang telah dilakukan pada tahap *preprocessing*. Masukkan file .xls dari masing-masing objek penelitian, lazada, bukalapak, dan blibli.



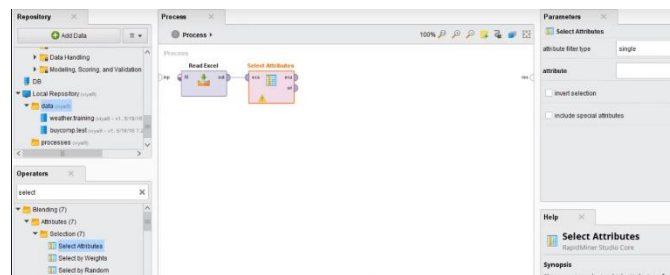
Gambar 3.11 naive bayes read excel

2. Kemudian, pilih *select attribute* untuk mengetahui atribut dari kata yang telah distemming.



Gambar 3.12 naive bayes sselect atributte.

3. Setelah itu, pilih metode *naive bayes* dalam menentukan akurasi data.



Gambar 3.13 naive bayes classification

4. Hasil dari proses ini adalah data akurasi dari file excel (.xls) dari ketiga situs belanja, yakni lazada, bukalapak dan blibli.
5. Hasil akurasi data pada proses klasifikasi *naive bayes* ini selanjutnya akan dilakukan pengujian dengan menggunakan *10-fold cross validation* menggunakan aplikasi Rapidminer.

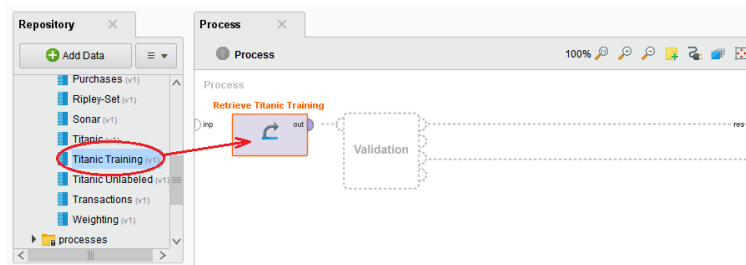
3.2.4 Analisis Validitas Data

Pada tahap analisis validitas data pengujian dilakukan dengan *10-fold cross validation*. Berikut adalah tahapan yang akan dilakukan dalam penghitungan menggunakan *cross validation* pada Rapidminer:

Validasi silang dilakukan dengan membagi contoh yang ditetapkan ke dalam bagian yang sama dan berputar melalui semua bagian, selalu menggunakan satu untuk pengujian dan yang lainnya untuk melatih model. Pada akhirnya, rata-rata semua akurasi pengujian disampaikan sebagai hasilnya. Ini merupakan cara yang bagus untuk menghitung keakuratan model dan harus menjadi pendekatan estimasi standar kita bila ada upaya penghitungan tambahan yang layak dilakukan.

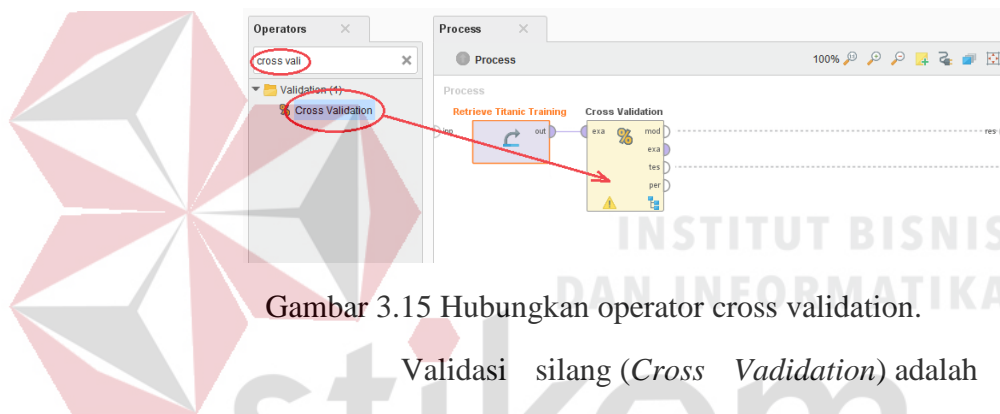
Langkah 1. Mengambil data dan memasukkannya ke dalam validasi.

1. Seret data *Titanic Training* ke dalam proses .



Gambar 3.14 Proses Training.

2. Tambahkan operator *Cross Validation* ke proses dan hubungkan.



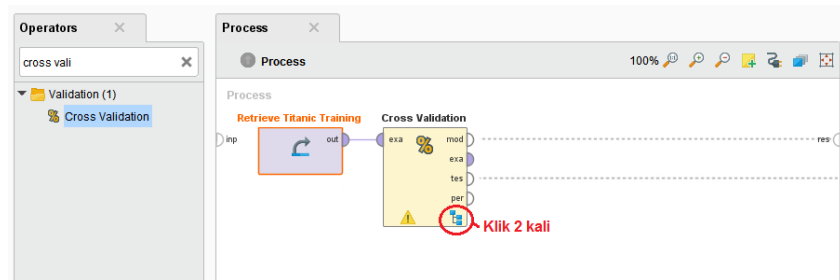
Gambar 3.15 Hubungkan operator cross validation.

Validasi silang (*Cross Validation*) adalah nama operator yang melakukan validasi silang dan perlu seperangkat contoh masukan yang berlabel. Secara default, Validasi silang membagi data menjadi 10 bagian yang berbeda, jadi kita menyebutnya validasi silang 10 kali lipat. Tentu saja Anda dapat mengubah jumlah lipatan di panel Parameter.

Langkah 2. Melatih dan menerapkan model.

Perhatikan ikon kecil di pojok kanan bawah *Cross Validation*. Klik dua kali pada operator untuk melihat sub-prosesnya.

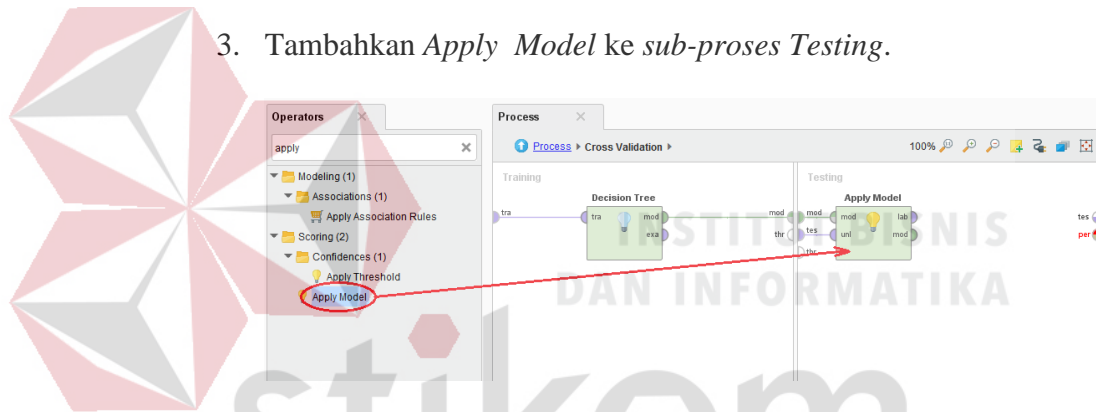
1. Klik dua kali operator *Cross Validation*. Panel Proses sekarang menunjukkan dua sub-proses bernama *Training* and *Testing*.



Gambar 3.16 Sub-proses training dan testing.

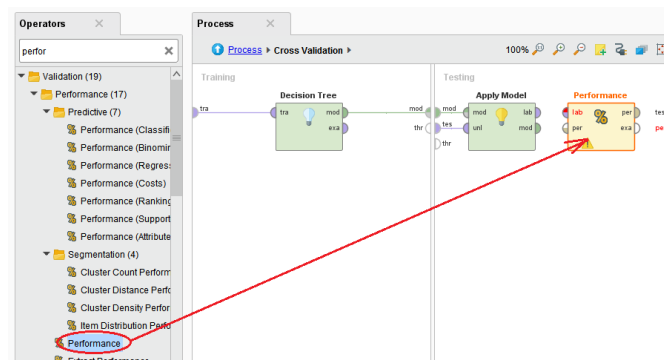
2. Sambungkan dengan *port* yang mengantarkan data pelatihan di sebelah kiri dan keluaran model di sebelah kanan.

3. Tambahkan *Apply Model* ke sub-proses *Testing*.



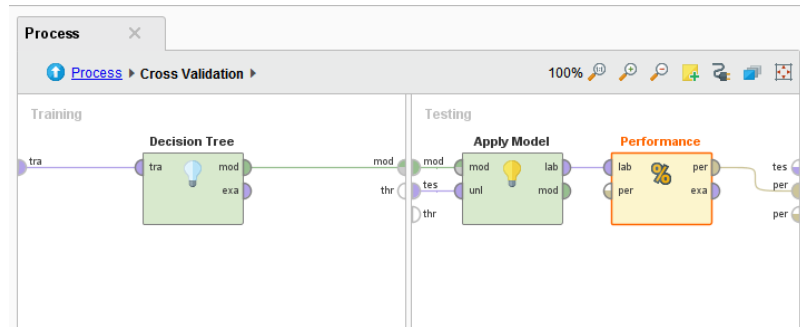
Gambar 3.17 Apply model sub-proses testing.

4. Hubungkan model dan data pengujian dari kiri ke *Apply Model*.
5. Juga tambahkan *Performance* ke *Testing*.



Gambar 3.18 Performance testing.

6. Hubungkan *output* dari *Apply Model* dengan Performance dan keluaran Performance dengan port rata-rata pertama "per" di sebelah kanan "per".

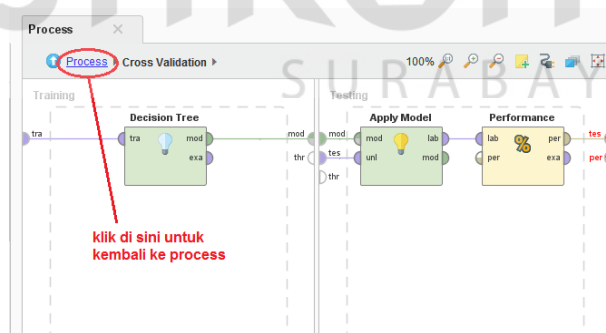


Gambar 3.19 Hubungkan output.

Setiap *performance* (kinerja) yang dikirimkan pada port "per" di sebelah kanan akan dirata-ratakan bersama-sama dan dikirim ke port output "per" dari operator *Cross Validation*.

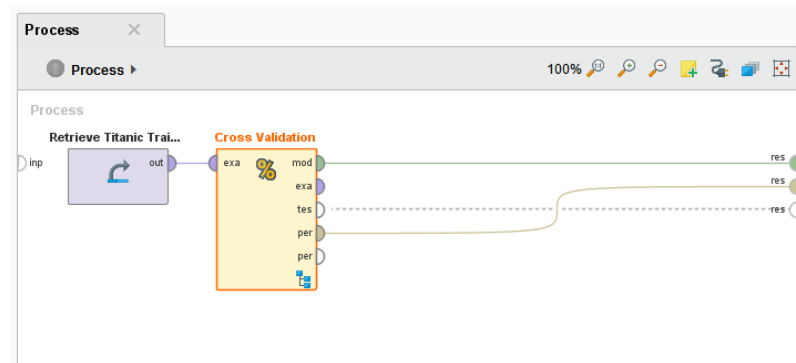
Langkah 3. Sambungkan ke port hasil dan jalankan proses.

1. Kembali ke proses utama dengan mengklik di sudut kiri atas panel Proses atau dengan mengklik link yang disebut "*Process*".



Gambar 3.20 Klik link proses.

2. Hubungkan model hijau dan port kinerja rata-rata kuning ke port hasil di sebelah kanan.



Gambar 3.21 Hubungkan port hijau dan kuning.

3. Jalankan proses dan periksa hasilnya.
4. Hasil dari proses pengujian ini adalah suatu file berbentuk .xls

yang kemudian akan diolah kembali pada proses visualisasi data word cloud menggunakan *Tagcrowd* untuk melihat kata yang paling banyak muncul. Data tersebut akan dijadikan sebagai bahan evaluasi tahap akhir.

3.2.5 Perbandingan Rating Secara Umum

Pada langkah berikut ini akan ditentukan rating dari seluruh kegiatan yang telah dilakukan mulai dari *crawling* data hingga testing yang telah dilakukan. Pada tahap ini akan dihitung seberapa besar presentase komentar positif, negatif dan netral yang didapatkan dari data *testing* pada aplikasi *RapidMiner*.

- a. Skala yang digunakan: 1 – 3.
- b. Masing-masing komponen penilaian dihitung persentase nilai positif dari total data training yang ada.
- c. Rating dihitung berdasarkan jumlah seluruh data training dibagi jumlah seluruh persentase nilai positif.
- d. Kemudian dibuat perangkingsannya.

Berikut adalah tabel yang akan digunakan untuk menentukan rating secara umum :

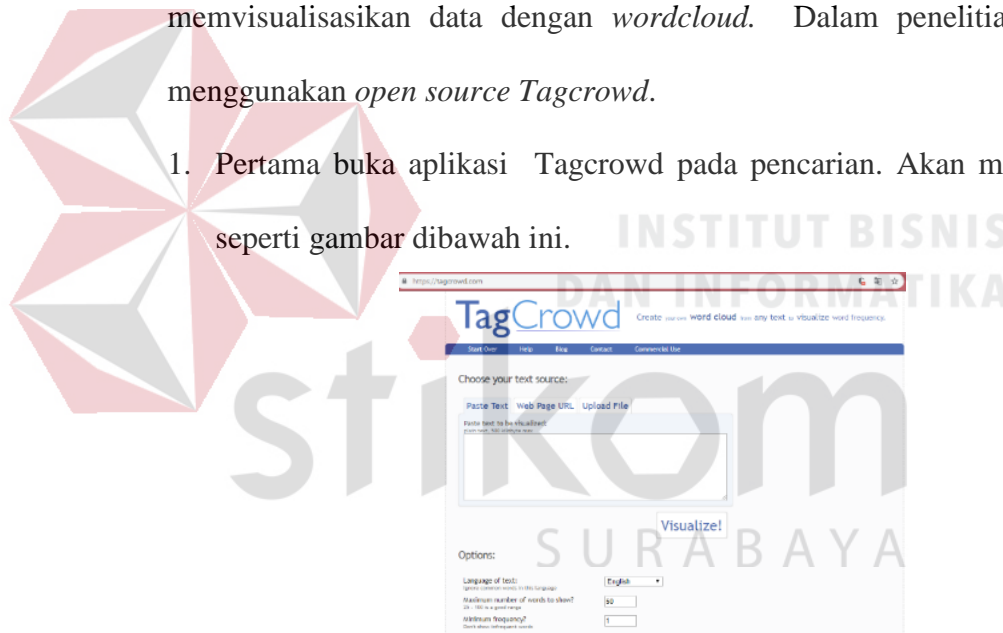
Tabel 3.3 perbandingan rating secara umum

Rank	Situs	Comment Count (cc)	Online Visitor (ov)	Rating
1				
2				
3				

3.2.6 Visualisasi Data dengan Word Cloud

Berikut adalah urutan langkah-langkah yang harus dilakukan dalam memvisualisasikan data dengan *wordcloud*. Dalam penelitian ini menggunakan *open source Tagcrowd*.

1. Pertama buka aplikasi Tagcrowd pada pencarian. Akan muncul seperti gambar dibawah ini.



Gambar 3.22 Tagcrowd visualisasi *wordcloud*.

2. Kemudian pilih upload file karena data yang akan divisualisasikan sudah terbentuk pada langkah-langkah sebelumnya, yakni pada tahap *stemming*.

The screenshot shows the TagCrowd website at the URL <https://tagcrowd.com>. The page has a blue header with the TagCrowd logo and navigation links: Start Over, Help, Blog, Contact, and Commercial Use. Below the header, there's a section titled "Choose your text source:" with three buttons: "Paste Text", "Web Page URL", and "Upload File". The "Upload File" button is selected. Below this, there's a file upload area with a text input field, a "Pilih File" button, and a message "Tidak ada file yang dipilih". A "Visualize!" button is located to the right of the upload area. Below the upload area, there's an "Options:" section with several settings:

- Language of text: English (dropdown menu)
- Ignore common words in this language: ☐ (checked)
- Maximum number of words to show?: 50 (input field)
- Minimum frequency?: 1 (input field)
- Don't show infrequent words: ☐ (checked)
- Show frequencies?: ☐ no (selected), ☐ yes
- Show word count next to each word: ☐ no (selected), ☐ yes
- Group similar words? (English only): ☐ no (selected), ☐ yes
- Convert to lowercase?: ☐ lowercase (selected), ☐ original

 At the bottom right, there is a "Download" button with a download icon.

Gambar 3.23 Tagcrowd Upload File.

3. Setelah memilih file yang dituju → klik *Visualize*, untuk melihat hasil dari *wordcloud*.
4. Hasil akhir dari proses visualisasi word cloud ini akan di jadikan bahan acuan dalam penelitian untuk menentukan peringkat popularitas dari lazada, bukalapak, maupun blibli sesuai dengan data dan pengujian yang telah dilakukan.
5. Kata-kata yang sering muncul dalam *wordcloud* akan di evaluasi untuk dijadikan solusi bagi pemilik situs belanja maupun konsumen.

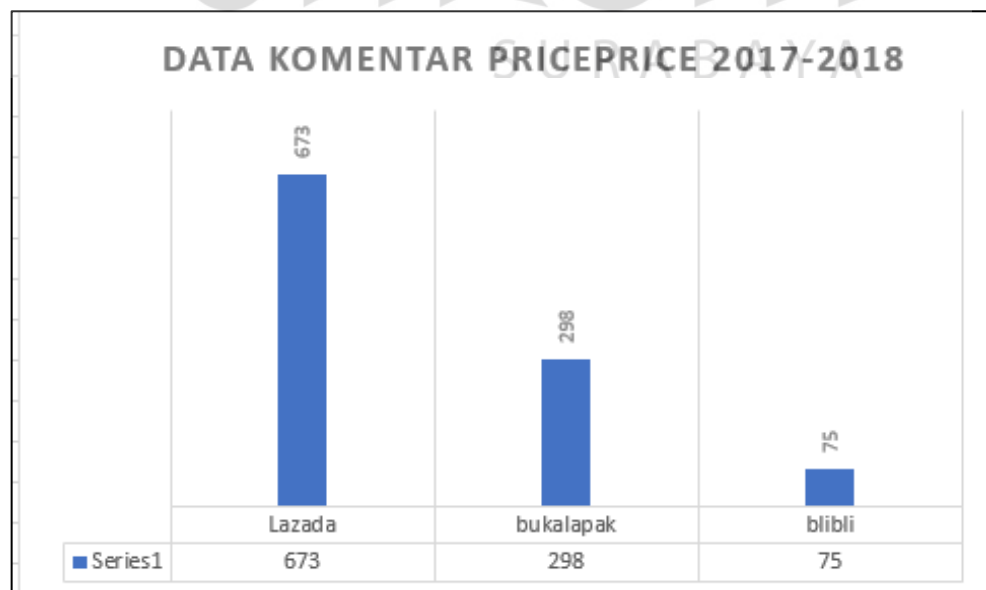
BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas hasil analisis berdasarkan pengolahan data pada platform website priceprice.com yang memuat komentar dari ketiga objek penelitian, yakni Lazada, Bukalapak dan Blibli.com yang telah dilakukan. Metode yang digunakan dalam klasifikasi teks komentar adalah *Naïve Bayes Classifier* (NBC).

4.1 Karakteristik Data dan Input Data

Website priceprice adalah salah satu bentuk layanan pelanggan khusus melalui *online* yang disediakan untuk menanggapi tanggapan, pendapat, kritik, saran dan masalah *complain*. Secara manual, dilakukan pengumpulan data terhadap jumlah *post* dari konsumen yang ditujukan kepada situs belanja *online* Lazada, Bukalapak dan Blibli.com. Gambar 4.1 menunjukkan jumlah komentar dari konsumen pada kurun waktu tahun 2017- 2018.



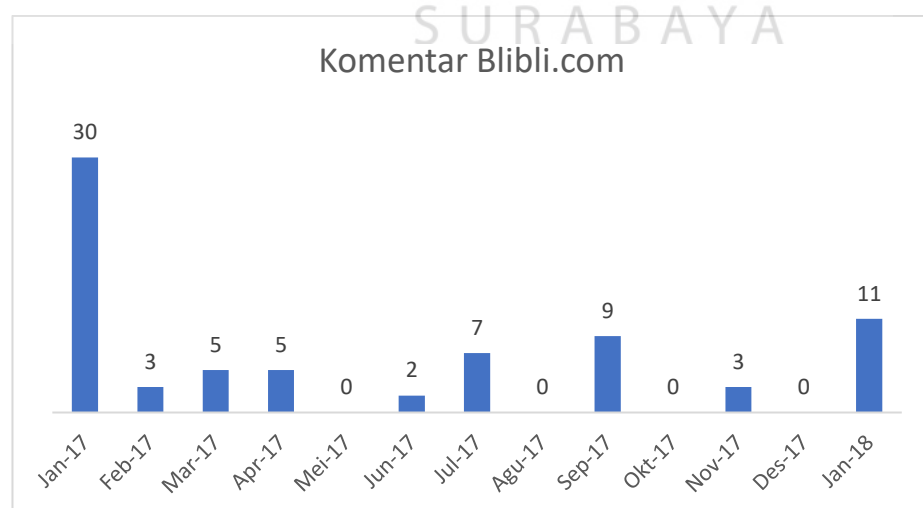
Gambar 4.1 Data komentar priceprice tahun 2017 – 2018.

Gambar 4.1 menunjukkan jumlah komentar konsumen pada tahun 2017 - 2018. Konsumen yang memberikan tanggapan, pendapat, kritik, saran dan masalah *complain* lebih banyak ditujukan pada akun Lazada dari pada akun Bukalapak dan Blibli.com. Dapat dilihat jumlah komentar yang ditujukan pada ketiga akun pada tahun 2017 - 2018, lazada menjadi situs yang paling banyak diulas dengan 673 komentar dalam kurun waktu 2 tahun.

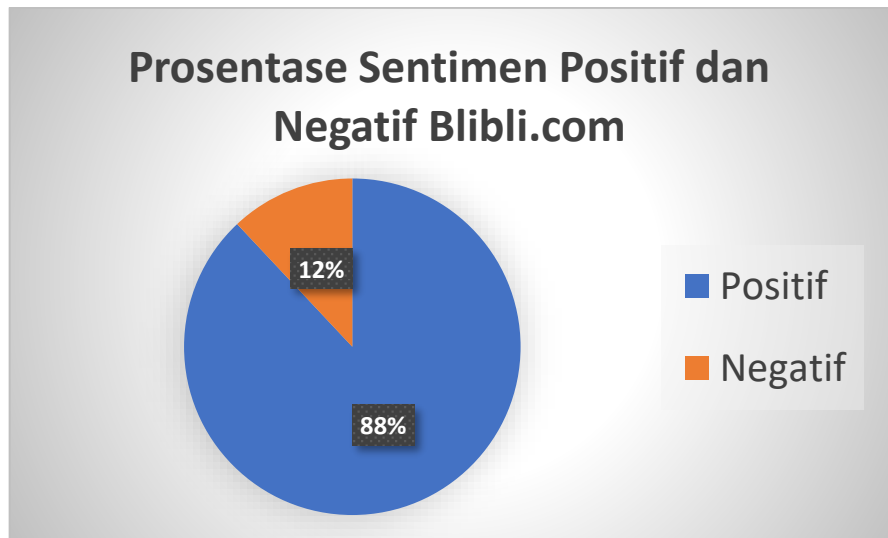
A. Situs Belanja online Blibli.com

Berikut merupakan prosentase komentar positif maupun negatif dari *online shop* Blibli.com berdasarkan perhitungan manual. Perhitungan manual dilakukan oleh peneliti untuk melakukan perbandingan dengan hasil perhitungan menggunakan aplikasi *Rapidminer* untuk menghasilkan data yang lebih akurat. Perhitungan ini dilihat berdasarkan komentar pengguna.

Apabila komentar yang ditulis mengandung kata-kata positif dan rating yang bagus, maka komentar tersebut masuk kedalam sentimen positif, begitu pula sebaliknya. Dapat dilihat pada lampiran.



Gambar 4.2 Perkembangan Komentar Blibli.com.



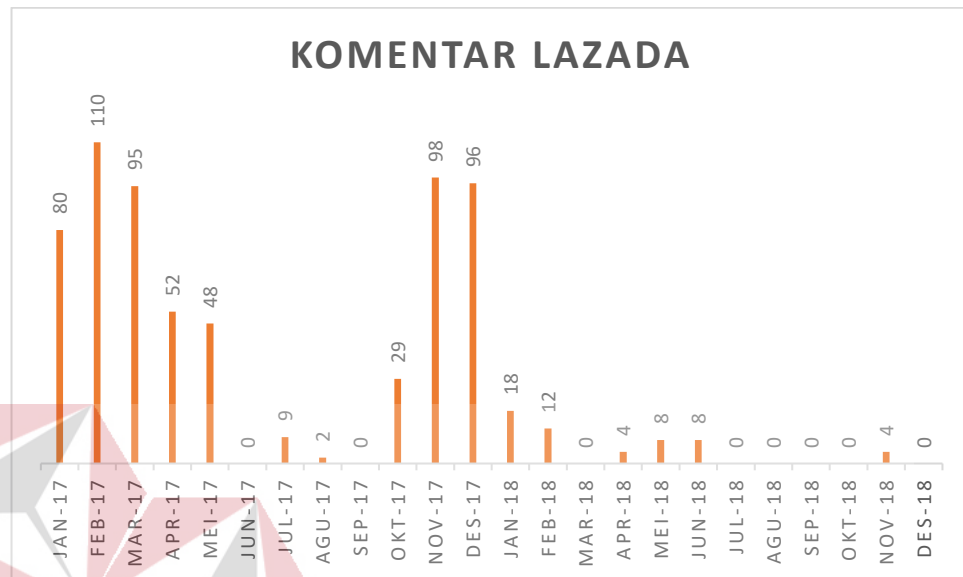
Gambar 4.3 Prosentase sentimen positif dan negatif Blibli.com.

Berdasarkan Gambar 4.2 dan gambar 4.3 dapat diketahui bahwa data yang digunakan untuk klasifikasi terdiri dari 75 komentar yang memuat sentimen negatif dan positif. Jumlah komentar positif yang dihitung secara manual pada *online shop* Blibli.com adalah sebanyak 66 dan 9 komentar negatif. Sentimen positif pada situs belanja online blibli.com memiliki prosentase yang lebih banyak jika dibandingkan dengan prosentase sentimen negatif. Prosentase sentimen negatif pada situs belanja online Blibli.com secara keseluruhan sebesar 12%, sedangkan prosentase sentimen positif adalah sebesar 88%.

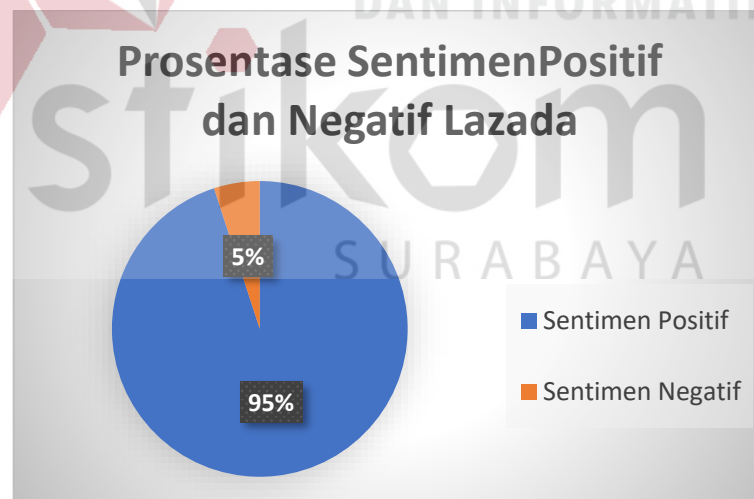
B. Situs Belanja online Lazada

Berikut merupakan prosentase komentar positif maupun negatif dari *online shop* Lazada berdasarkan perhitungan manual. Perhitungan manual dilakukan oleh peneliti untuk melakukan perbandingan dengan hasil perhitungan menggunakan aplikasi *Rapidminer* untuk menghasilkan data yang lebih akurat. Perhitungan ini dilihat berdasarkan komentar pengguna.

Apabila komentar yang ditulis mengandung kata-kata positif dan rating yang bagus, maka komentar tersebut masuk kedalam sentimen positif, begitu pula sebaliknya. Dapat dilihat pada lampiran.



Gambar 4.4 Perkembangan Komentar Lazada.



Gambar 4.5 Prosentase sentimen positif dan negatif Lazada.

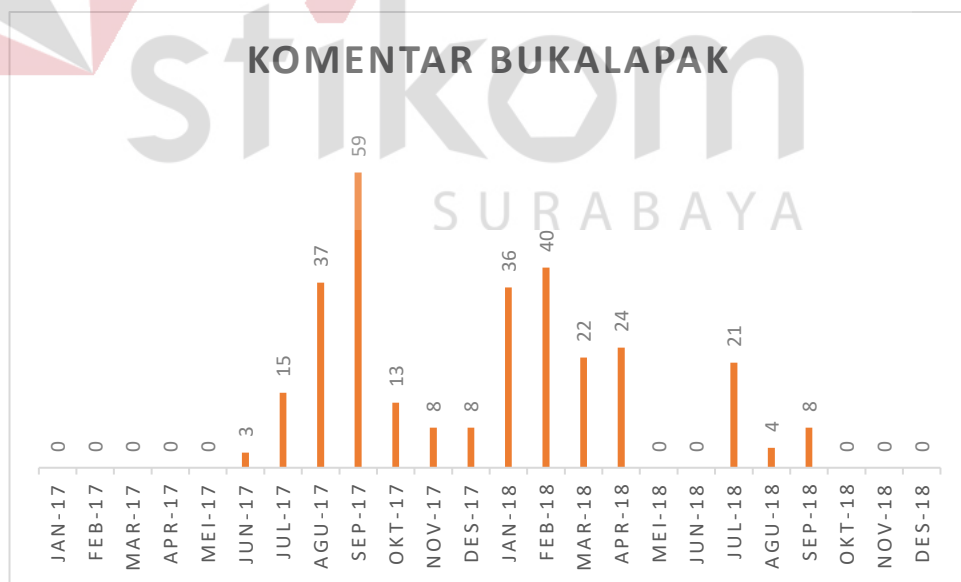
Berdasarkan Gambar 4.4 dan gambar 4.5 dapat diketahui bahwa data yang digunakan untuk klasifikasi terdiri dari 673 komentar yang memuat sentimen negatif dan positif. Jumlah komentar positif yang dihitung secara manual pada *online shop* Lazada adalah sebanyak 640 dan 33 komentar negatif. Sentimen

positif pada situs belanja *online* Lazada memiliki prosentase yang lebih banyak jika dibandingkan dengan prosentase sentimen negatif. Prosentase sentimen negatif pada situs belanja *online* Lazada secara keseluruhan sebesar 5%, sedangkan prosentase sentimen positif adalah sebesar 95%.

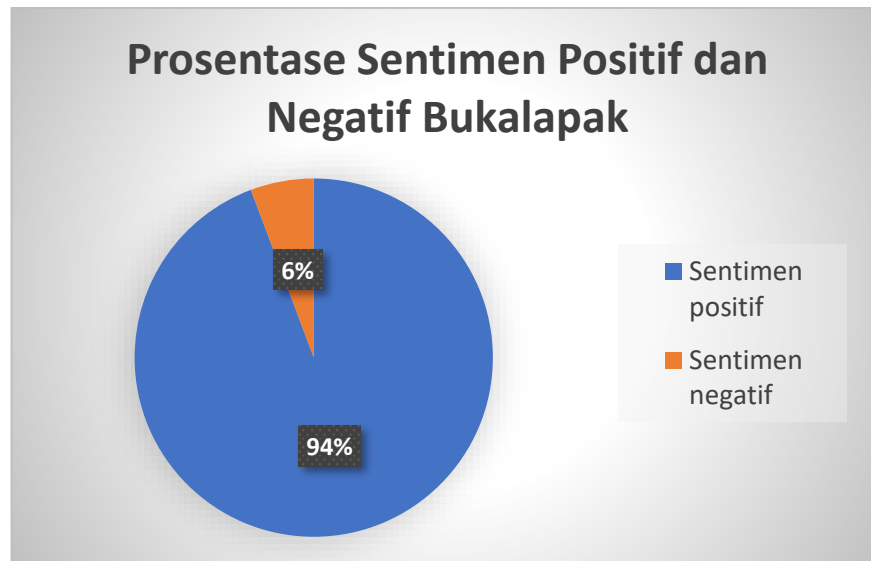
C. Situs Belanja online Bukalapak

Berikut merupakan prosentase komentar positif maupun negatif dari *online shop* Bukalapak berdasarkan perhitungan manual. Perhitungan manual dilakukan oleh peneliti untuk melakukan perbandingan dengan hasil perhitungan menggunakan aplikasi *Rapidminer* untuk menghasilkan data yang lebih akurat. Perhitungan ini dilihat berdasarkan komentar pengguna.

Apabila komentar yang ditulis mengandung kata-kata positif dan rating yang bagus, maka komentar tersebut masuk kedalam sentimen positif, begitu pula sebaliknya. Dapat dilihat pada lampiran.



Gambar 4.6 Perkembangan Komentar Lazada.



Gambar 4.7 Prosentase sentimen positif dan negatif Lazada.

Berdasarkan Gambar 4.6 dan gambar 4.7 dapat diketahui bahwa data yang digunakan untuk klasifikasi terdiri dari 298 komentar yang memuat sentimen negatif dan positif. Jumlah komentar positif yang dihitung secara manual pada *online shop* Bukalapak adalah sebanyak 281 dan 17 komentar negatif. Sentimen positif pada situs belanja online Bukalapak memiliki prosentase yang lebih banyak jika dibandingkan dengan prosentase sentimen negatif. Prosentase sentimen negatif pada situs belanja online Bukalapak secara keseluruhan sebesar 6%, sedangkan prosentase sentimen positif adalah sebesar 94%.

Berikut ditampilkan sentimen analisis studi kasus ini dari ketiga situs belanja *online* berdasarkan perhitungan manual. Data dibawah ini adalah data yang akan dijadikan sebagai pembandingan untuk mengukur akurasi pada tahap berikutnya. Dapat dilihat pada Tabel 4.1

Tabel 4.1 Keseluruhan Sentimen Positif dan Negatif.

No.	Situs Belanja	Jumlah Sentimen (manual)		Prosentase Sentimen (manual)	
		Positif	Negatif	Positif	Negatif
1	Lazada	640	33	95%	5%
2	Bukalapak	281	17	94%	6%
3	Blibli.com	66	9	88%	12%

4.2 Data Preprocessing

Sebelum dilakukan analisis, langkah selanjutnya yang akan dilakukan adalah pra proses pada data komentar *platform website priceprice.com* yang memuat komentar ketiga situs belanja online, yakni lazada, bukalapak, dan blibli.com terdiri dari beberapa tahapan, yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming*. Proses *case folding* adalah mengubah data teks menjadi lowercase dengan tujuan agar kata yang sama namun berbeda secara penulisan huruf kapital dan tidak, tidak dianggap kata yang berbeda. Contoh data sebelum dan sesudah melalui tahap lowercase ditampilkan pada Tabel 4.2

Tabel 4.2 Proses case folding.

Sebelum proses <i>Case Folding</i>	Setelah Proses <i>Case Folding</i>
Belanja di Blibli.com emang mudah banget, ditambah produk yg dijual cukup murah tersedia pilihan pembayaran ditempat (COD) membantu bgt buat pemula yg takut	belanja di blibli.com emang mudah banget, ditambah produk yg dijual cukup murah tersedia pilihan pembayaran ditempat (cod) membantu bgt buat pemula yg takut barang tidak

Sebelum proses <i>Case Folding</i>	Setelah Proses <i>Case Folding</i>
barang tidak diterima. Overall , semua pelayanan udah bagus..	diterima. overall , semua pelayanan udah bagus..
.	.
.	.
.	.
Awalnya saya senang pesan di lazada, sudah lebih dari 2 tahun saya menjadi pelanggan lazada... tapi 1 tahun belakangan ini saya kecewa banget karena penangannya sangat mengecewakan...	awalnya saya senang pesan di lazada, sudah lebih dari 2 tahun saya menjadi pelanggan lazada... tapi 1 tahun belakangan ini saya kecewa banget karena penangannya sangat mengecewakan...
.	.
.	.
.	.
Sumpah ya kesel banget.. Beli token pulsa dari tgl 15 sampe tanggal segini belum dikirim kode tokeny.. Malah statusnya kedaluwarsa.. Padaahal saya udah bayar dn kirim bukti transfer.. Wktu Hari minggu bilanganya ga bsa mutasi rekening bank.. Giliran ini udah Hari senin suruh tunggu.. Tunggu apaan?? Mana tanggung jawab buka lapak ni. Bukti transfer udh dikirim jga ga bsa mroses.	sumpah ya kesel banget.. beli token pulsa dari tgl 15 sampe tanggal segini belum dikirim kode tokeny.. malah statusnya kedaluwarsa.. padahal saya udah bayar dn kirim bukti transfer.. wktu hari minggu bilanganya ga bsa mutasi rekening bank.. giliran ini udah hari senin suruh tunggu.. tunggu apaan?? mana tanggung jawab buka lapak ni. bukti transfer udh dikirim jga ga bsa mroses.

Setelah melakukan proses *case folding* pada data teks, langkah selanjutnya adalah proses *tokenizing* dimana proses memecah kalimat menjadi kata-kata. Contoh data sebelum dan sesudah melalui tahap *stopword* dan *tokenizing* ditampilkan pada Tabel 4.3

Tabel 4.3 Proses Tokenizing.

Sebelum proses <i>Tokenizing</i>	Setelah Proses <i>Tokenizing</i>
belanja di blibli.com emang mudah banget, ditambah produk yg dijual cukup murah tersedia pilihan pembayaran ditempat (cod) membantu bgt buat pemula yg takut barang tidak diterima . overall , semua pelayanan udah bagus..	belanja di blibli com emang mudah banget, di tambah produk yg di jual cukup murah tersedia pilihan pembayaran di tempat (cod) membantu bgt buat pemula yg takut barang tidak di terima . over all , semua pelayanan udah bagus...
awalnya saya senang pesan di lazada, sudah lebih dari 2 tahun saya menjadi pelanggan lazada... tapi 1 tahun belakangan ini saya kecewa banget karena penangannya sangat mengecewakan...	awalnya saya senang pesan di lazada, sudah lebih dari 2 tahun saya menjadi pelanggan lazada... tapi 1 tahun belakangan ini saya kecewa banget karena penangannya sangat mengecewakan...
sumpah ya kesel banget.. beli token pulsa dari tgl 15 sampe tanggal segini belum dikirim kode tokenny.. malah statusnya kedaluwarsa.. padahal saya udah bayar dn kirim bukti transfer.. wktu hari minggu bilanganya ga bsa mutasi rekening bank.. giliran ini udah hari senin	sumpah ya kesel banget.. beli token pulsa dari tgl 15 sampe tanggal segini belum dikirim kode tokenny.. malah statusnya kedaluwarsa.. padahal saya udah bayar dn kirim bukti transfer.. wktu hari minggu bilanganya ga bsa mutasi rekening bank.. giliran ini udah hari senin suruh tunggu.. tunggu

Sebelum proses <i>Tokenizing</i>	Setelah Proses <i>Tokenizing</i>
suruh tunggu.. tunggu apaan?? mana tanggung jawab buka lapak ni. bukti transfer udh dikirim jga ga bsa mroses.	apaan?? mana tanggung jawab buka lapak ni. bukti transfer udh di kirim jga ga bsa mroses.

Setelah melakukan proses *tokenizing* pada data teks, langkah selanjutnya adalah proses *filtering*. *Filtering* pada data komentar merupakan proses membersihkan komentar dari kata yang tidak diperlukan untuk mengurangi noise. Kata yang dihilangkan dalam komentar adalah karakter HTML, *emoticons*, *hashtag* (#), *username* (@username), URL, dan menghilangkan *stopword*. Contoh data sebelum dan sesudah melalui tahap *filtering* ditampilkan pada Tabel 4.4

Tabel 4.4 Proses *Filtering*.

Sebelum proses <i>Filtering</i>	Setelah Proses <i>Filtering</i>
belanja di blibli com emang mudah banget, di tambah produk yg di jual cukup murah tersedia pilihan pembayaran di tempat (cod) membantu bgt buat pemula yg takut barang tidak di terima. over all, semua pelayanan udah bagus...	Belanja blibli mudah banget tambah produk jual cukup murah sedia pilih bayar tempat cod bantu pemula takut barang tidak terima over all semua layanan bagus
.	.
.	.
.	.
awalnya saya senang pesan di lazada, sudah lebih dari 2 tahun saya menjadi pelanggan lazada... tapi 1 tahun belakangan ini saya kecewa banget karena penangannya sangat mengecewakan...	awal saya senang pesan lazada sudah lebih dari tahun saya jadi pelanggan lazada tahun belakang kecewa banget penanganan sangat kecewa
.	.

Sebelum proses <i>Filtering</i>	Setelah Proses <i>Filtering</i>
---------------------------------	---------------------------------

sumpah ya kesel banget.. beli token pulsa dari tgl 15 sampe tanggal segini belum dikirim kode tokenny.. malah statusnya kedaluwarsa.. padahal saya udah bayar dn kirim bukti transfer.. wktu hari minggu bilangny ga bsa mutasi rekening bank.. giliran ini udah hari senin suruh tunggu.. tunggu apaan?? mana tanggung jawab buka lapak ni. bukti transfer udh di kirim jga ga bsa mproses.

sumpah kesel banget beli token pulsa dari tanggal gini belum kirim kode token status kedaluwarsa padahal saya udah bayar kirim bukti transfer hari minggu bilang mutasi rekening bank giliran hari senin suruh tunggu tunggu apa mana tanggung jawab bukalahap bukti transfer kirim

Proses stemming dilakukan untuk mendapatkan kata dasar dengan cara menghilangkan awalan, akhiran, sisipan, dan *confixes* (kombinasi dari awalan dan akhiran). Pada penelitian ini kamus stemming yang digunakan adalah *Liu's word list*. Berikut adalah contoh data sebelum dan sesudah melalui tahap *stemming* ditampilkan pada Tabel 4.5

Tabel 4.5 Proses Stemming.

Sebelum proses <i>Stemming</i>	Setelah Proses <i>Stemming</i>
Belanja blibli mudah banget tambah produk jual cukup murah sedia pilih bayar tempat cod bantu pemula takut barang tidak terima over all semua layanan bagus	'belanja' 'blibli' 'mudah' 'banget' 'tambah' 'produk' 'jual' 'cukup' 'murah' 'sedia' 'pilih' 'bayar' 'tempat' 'bantu' 'pemula' 'takut' 'barang' 'tidak' 'terima' 'over' 'all' 'semua' 'layanan' 'bagus'
.	.
.	.
.	.

awal saya senang pesan lazada sudah
lebih dari tahun saya jadi pelanggan
lazada tahun belakang kecewa
banget penanganan sangat kecewa

'awal' 'saya' 'senang' 'pesan' 'lazada'
'sudah' 'lebih' 'dari' 'tahun' 'saya'
'jadi' 'pelanggan' 'lazada' 'tahun'
'belakang' 'kecewa' 'banget'
'penanganan' 'sangat' 'kecewa'

.

.

.

.

.

.

sumpah kesel banget beli token pulsa
dari tanggal gini belum kirim kode
token status kedaluwarsa padahal
saya udah bayar kirim bukti transfer
hari minggu bilang mutasi rekening
bank giliran hari senin suruh tunggu
tunggu apa mana tanggung jawab
bukalapak bukti transfer kirim

'sumpah' 'kesel' 'banget' 'beli' 'token'
'pulsa' 'dari' 'tanggal' 'gini' 'belum'
'kirim' 'kode' 'token' 'status'
'kedaluwarsa' 'padahal' 'saya' 'udah'
'bayar' 'kirim' 'bukti' 'transfer' 'hari'
'minggu' 'bilang' 'mutasi' 'rekening'
'bank' 'giliran' 'hari' 'senin' 'suruh'
'tunggu' 'tunggu' 'apa' 'mana'
'tanggung' 'jawab' 'bukalapak' 'bukti'
'transfer' 'kirim'

Hasil dari proses stemming dipakai sebagai kata kunci dari data komentar yang akan dianalisis dengan menggunakan metode NBC (*Naive Bayes Classification*). Tabel 4.6, Tabel 4.7 dan Tabel 4.8 menunjukkan perhitungan frekuensi tiap kata kunci pada situs belanja *online*.

Tabel 4.6 Frekuensi data setelah proses *stemming* Lazada.

Attribute Name	Total Occurences
lazada	596
barang	171
belanja	154
pelayanan	107
puas	91
kasih	89
online	87
semoga	86
harga	80
cepat	72
murah	70
beli	66
berbelanja	57
diskon	57
suka	47

Berdasarkan Tabel 4.6 menunjukkan perhitungan frekuensi data setelah proses *stemming* pada *online shop* Lazada terdapat beberapa kata kunci. Pada komentar pertama didapatkan bahwa kata “lazada” disebutkan pada sebanyak 596 kali, disusul dengan kata “barang” dengan 171 kali, kata “belanja” sebanyak 154 kali, dan seterusnya.

Tabel 4.7 Frekuensi data setelah proses *stemming* Bukalapak.

Attribute Name	Total Occurences
bukalapak	275
belanja	83
barang	72
kasih	45
terima	45
online	44
banget	35
beli	35
pembeli	31
semoga	26
sukses	24
aman	23
harga	23
berbelanja	22
buka	21

Berdasarkan Tabel 4.7 menunjukkan perhitungan frekuensi data setelah proses *stemming* pada *online shop* Bukalapak terdapat beberapa kata kunci. Pada komentar didapatkan bahwa kata “bukalapak” disebutkan 275 kali, disusul dengan kata “belanja” yang disebutkan sebanyak 83 kali, kata “barang” disebutkan sebanyak 72 kali, dan seterusnya.

Tabel 4.8 Frekuensi data setelah proses *stemming* Blibli.com.

Attribute Name	Total Occurences
blibli	55
barang	34
aman	22
mudah	19
puas	18
belanja	17
pembayaran	16
cepat	14
voucher	10
murah	9
online	9
bagus	8
customer	8
pengirimannya	8
membeli	7



INSTITUT BISNIS
DAN INFORMATIKA

stikom

SURABAYA

Berdasarkan Tabel 4.8 menunjukkan perhitungan frekuensi data setelah proses *stemming* pada *online shop* Blibli.com terdapat beberapa kata kunci. Pada komentar didapatkan bahwa kata “blibli” disebutkan 55 kali, disusul dengan kata “barang” yang disebutkan sebanyak 34 kali, kata “aman” disebutkan sebanyak 22 kali, dan seterusnya.

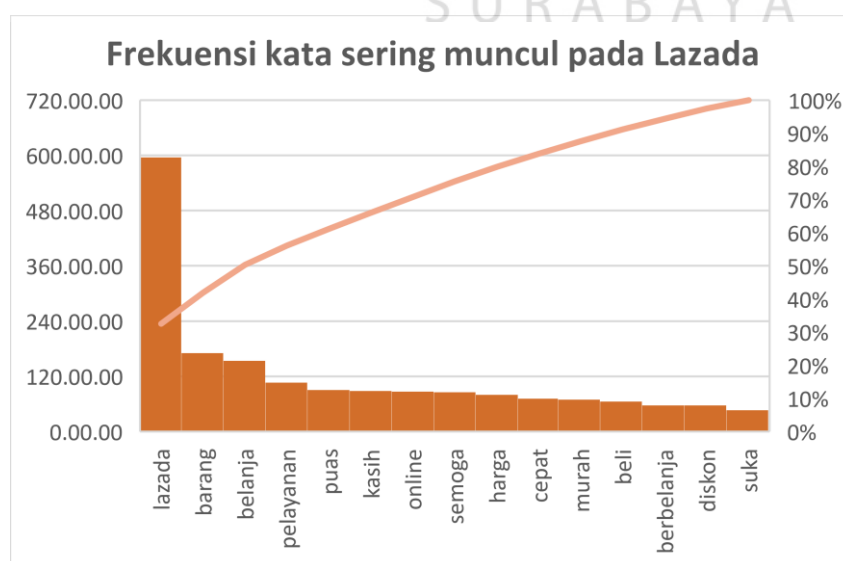
Setelah dilakukan *pre-processing* pada data teks, jumlah kata kunci dapat diketahui pada situs belanja online lazada, bukalapak dan blibli.com. Hasil dari proses *stopword* dan *tokenizing* dipakai sebagai kata kunci dari data komentar yang akan dianalisis dengan menggunakan metode NBC. Tabel 4.9 menunjukkan

perhitungan frekuensi tiap kata kunci pada *online shop* Bukalapak, Blibli.com dan Lazada.

Tabel 4.9 *Count Vectorized* untuk Komentar

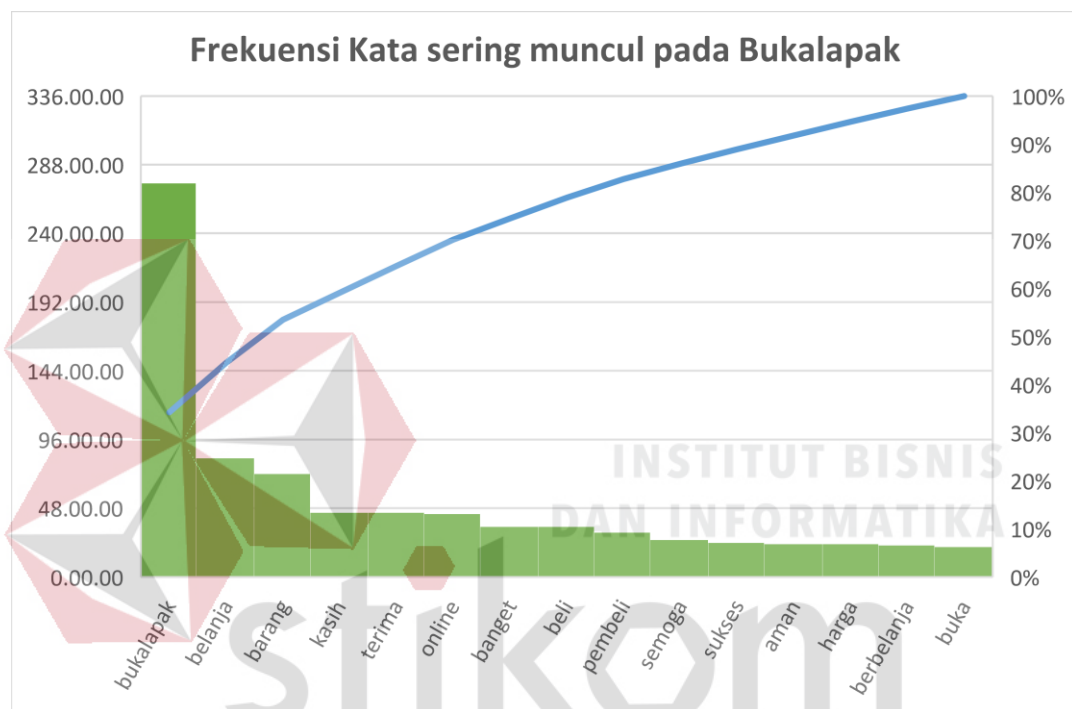
Komentar	Sentimen	Kata Kunci					
		aamiin	abis	acungi	wiraswasta	web
1	1	0	0	0	0	0
2	0	0	1	0	1	0
3	1	1	0	1	0	1
.
n	0	1	1	0	-----	0	0

Frekuensi setiap kata dalam komentar lazada, bukalapak dan blibli.com ditunjukkan pada Gambar 4.8, Gambar 4.9 dan Gambar 4.10.



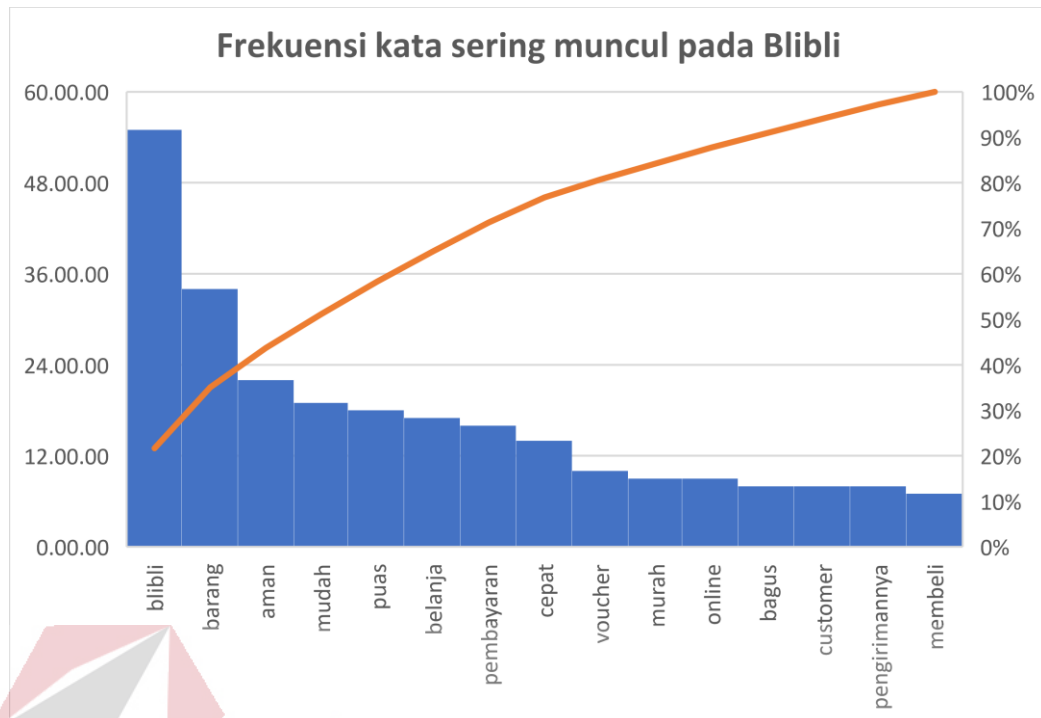
Gambar 4.8 Frekuensi data komentar lazada

Berdasarkan Gambar 4.8 dapat diketahui frekuensi dari 15 kata tertinggi situs belanja *online* Lazada. Kata “barang” merupakan kata yang paling banyak dibicarakan oleh konsumen. Hal tersebut dikarenakan bahwa Lazada adalah *market place* yang banyak menjual barang daripada jasa. Kata “bayar” merupakan kata kedua yang paling banyak dibicarakan.



Gambar 4.9 Frekuensi data komentar bukalapak

Berdasarkan Gambar 4.9 dapat diketahui frekuensi dari 15 kata tertinggi akun bukalapak. Kata “bukalapak” merupakan kata yang paling banyak dibicarakan oleh konsumen Bukalapak, sedangkan kata “belanja” merupakan kata yang paling banyak kedua dibicarakan oleh konsumen. Disusul dengan kata “barang” pada posisi ketiga, hal tersebut dikarenakan bahwa Bukalapak adalah *market place* yang banyak menjual barang daripada jasa.



Gambar 4.10 Frekuensi data komentar blibli.

Berdasarkan Gambar 4.10 dapat diketahui frekuensi dari 15 kata tertinggi *online shop* Blibli.com. Kata “blibli” merupakan kata yang paling banyak dibicarakan oleh konsumen Blibli, sedangkan kata “barang” merupakan kata yang paling banyak kedua dibicarakan oleh konsumen. Disusul dengan kata “aman” pada posisi ketiga, hal tersebut dikarenakan keamanan sangat dijunjung tinggi oleh *market place* ini.

Data diatas adalah hasil stemming kata positif yang terkandung dalam masing-masing komentar pengguna *website priceprice.com*. Berikut adalah frekuensi kata negatif yang diperoleh dalam penelitian ini dari hasil *stemming*. Dapat dilihat pada Tabel 4.10, Tabel 4.11 dan Tabel 4.12.

Tabel 4.10 Frekuensi kata Negatif Blibli.com.

Attribute Name	Total O...
blibli	6
bli	4
pengirimannya	4
lambat	3
barang	2
barangnya	2
beli	2
bertanggung	2
com	2
customer	2
daerah	2
dikirim	2
diperbaiki	2
diperjalan	2
ekspedisi	2

Berdasarkan Tabel 4.10 dapat diketahui frekuensi dari 10 kata tertinggi *online shop* Blibli.com. Kata “blibli” merupakan kata yang paling banyak dibicarakan oleh konsumen Blibli, sedangkan kata “pengiriman” merupakan kata yang paling banyak ketiga dibicarakan oleh konsumen setelah kata “bli”. Yang dimaksudkan disini adalah pengiriman pada *online shop* Blibli.com kurang dapat diterima oleh konsumen.

Tabel 4.11 Frekuensi kata Negatif Bukalapak.

Attribute Name	Total Occurences
kebingungan	1
kedua	1
kemajuan	1
kemudian	1
kena	1
keperluan	1
kesanku	1
kesekian	1
keseringan	1
ketersediaan	1

Berdasarkan Tabel 4.11 dapat diketahui frekuensi dari 10 kata tertinggi *online shop* Bukalapak. Kata “kebingungan” merupakan kata yang paling banyak dibicarakan oleh konsumen Blibli, sedangkan kata “kedua” merupakan kata yang paling banyak kedua dibicarakan oleh konsumen. Mungkin kebingungan yang dimaksud disini adalah kebingungan dalam aplikasi karena kata berikutnya yang muncul adalah “kemajuan”.

Tabel 4.12 Frekuensi kata Negatif Lazada.

Atribut	Total Occurences
cacat	20
buruk	16
bintang	12
berusaha	9
bersaing	7
biaya	6
bonus	6
beserta	6
browsing	4
bertransaksi	4

Berdasarkan Tabel 4.12 dapat diketahui frekuensi dari 10 kata tertinggi *online shop* Lazada. Kata “susah” merupakan kata yang paling banyak dibicarakan oleh konsumen Lazada, sedangkan kata “tokonya” merupakan kata yang paling banyak kedua dibicarakan oleh konsumen. Mungkin susah yang dimaksud disini adalah susah dalam komunikasi dengan toko yang dituju karena kata berikutnya yang muncul adalah “tokonya”.

4.3 Klasifikasi Naive Bayes (NBC)

Bayesian classification merupakan metode pengklasifikasian statistik yang dapat digunakan untuk memprediksi peluang keanggotaan suatu kelas. Langkah pertama dalam mengklasifikasikan data komentar adalah menguji model menggunakan data training. Model yang telah diuji dengan data training kemudian digunakan untuk mengklasifikasikan data testing ke dalam dua kelas sentimen yaitu positif dan negatif.

A. Klasifikasi *Naive Bayes* pada *Online Shop Lazada*.

Untuk menentukan suatu komentar masuk ke dalam klas negatif (0) atau kelas positif (1) dapat dilihat dari probabilitas berikut $P(X|C_i)P(C_i)$. Model *Naive Bayes Classifier* (NBC) yang digunakan dalam penelitian ini adalah sebagai berikut.

Untuk komentar pertama setelah proses *stemming* : “kiriman, cepat, rapi, barang, sesuai, foto, agak, besar”.

Tabel 4.13 NBC Lazada

Kata	Positif	Negatif
Kiriman	4	1
Cepat	80	2
Rapi	11	1
Barang	173	2
Sesuai	45	1
Foto	4	1
Agak	2	1
Total	317	8

$$\begin{aligned}
 \text{Class positif (1)} &= \frac{4}{5} \times \frac{80}{82} \times \frac{11}{12} \times \frac{173}{175} \times \frac{45}{46} \times \frac{4}{5} \times \frac{2}{3} \times \frac{640}{673} \\
 &= \frac{21922560}{594090000} \times \frac{640}{673} \\
 &= \frac{14030438400}{399822570000} \\
 &= 0,035090
 \end{aligned}$$

$$\begin{aligned}
 \text{Class negatif (0)} &= \frac{1}{5} \times \frac{2}{82} \times \frac{1}{12} \times \frac{2}{175} \times \frac{1}{46} \times \frac{1}{5} \times \frac{1}{3} \times \frac{33}{673} \\
 &= \frac{4}{594090000} \times \frac{33}{673} \\
 &= \frac{132}{399822570000} \\
 &= 0,000334
 \end{aligned}$$

Naïve bayes classifier (Normalisasi):

$$\begin{aligned}
 P(X|C_0)P(C_0) &= \frac{\text{Jumlah (class=0)}}{(\text{Jumlah (class=0)} + \text{Jumlah (class=1)})} \\
 &= \frac{0,000334}{0,000334 + 0,035090} \\
 &= 0,0094
 \end{aligned}$$

$$\begin{aligned}
 P(X|C_1)P(C_1) &= \frac{\text{Jumlah (class=1)}}{(\text{Jumlah (class=0)} + \text{Jumlah (class=1)})} \\
 &= \frac{0,035090}{0,000334 + 0,035090} \\
 &= 0,6469
 \end{aligned}$$

Keputusan : “**positif**”.

Tabel 4.14 Probabilitas Klasifikasi NBC pada Lazada.

Testing Komentar	Probabilitas Negatif	Probabilitas Positif	Keputusan
1	0,0094	0,6469	Positif
2	0,0580	0, 6429	Positif
3	0,0120	0,6888	Positif
4	0,0154	0,7851	Positif
5	0,8910	0,1094	Negatif
6	0,2987	0,7025	Positif
7	0,2475	0,7539	Positif
8	0,4577	0,5430	Positif
9	0,1780	0,8220	Positif
.	.	.	.
.	.	.	.
.	.	.	.
673	0,4043	0,5961	Positif

Berdasarkan Tabel 4.14 dapat diketahui probabilitas klasifikasi dengan metode *Naïve Bayes Classifier* (NBC) pada *online Shop* Lazada. Jika probabilitas sentimen positif lebih tinggi jika dibandingkan dengan probabilitas sentimen negatif, maka akan memberikan hasil keputusan positif begitupula sebaliknya. Pada data komentar yang pertama, probabilitas sentimen negatif adalah sebesar 0,0094, sedangkan probabilitas sentimen positif adalah sebesar 0,6469. Keputusan yang dapat diambil adalah kalimat pertama masuk kedalam sentimen positif, dikarenakan probabilitas sentimen positif lebih besar jika dibandingkan dengan probabilitas sentimen negatif.

B. Klasifikasi *Naive Bayes* pada Online Shop Bukalapak.

Untuk menentukan suatu komentar masuk ke klas negatif (0) atau kelas positif (1) dapat dilihat dari probabilitas berikut $P(X|C_i)P(C_i)$. Model *Naïve Bayes Classifier* (NBC) yang digunakan dalam penelitian ini adalah sebagai berikut.

Komentar pertama setelah proses *stemming* : “kesel, beli, belum, kirim, kode, tidak, kadaluwarsa, kecewa”.

Tabel 4.15 NBC Bukalapak.

Kata	Positif	Negatif
Kesel	1	4
Beli	2	38
Belum	1	2
Kirim	1	6
Kode	1	6
Tidak	3	1
Kadaluwarsa	2	2
Kecewa	10	1
Total	22	58

$$\begin{aligned}
 \text{Class positif (1)} &= \frac{1}{5} \times \frac{2}{40} \times \frac{1}{3} \times \frac{1}{7} \times \frac{1}{7} \times \frac{1}{4} \times \frac{2}{4} \times \frac{10}{11} \times \frac{281}{298} \\
 &= \frac{40}{5174400} \times \frac{281}{298} \\
 &= \frac{11240}{1541971200} \\
 &= 0,000728
 \end{aligned}$$

$$\begin{aligned}
 \text{Class negatif (0)} &= \frac{4}{5} \times \frac{38}{40} \times \frac{2}{3} \times \frac{6}{7} \times \frac{6}{7} \times \frac{1}{4} \times \frac{2}{4} \times \frac{1}{11} \times \frac{17}{298} \\
 &= \frac{21888}{5174400} \times \frac{17}{298} \\
 &= \frac{372096}{1541971200} \\
 &= 0,024131
 \end{aligned}$$

Naive bayes classifier (normalisasi):

$$P(X|C_0)P(C_0) = \frac{\text{Jumlah (class=0)}}{(\text{Jumlah (class=0)} + \text{Jumlah (class=1)})}$$

$$= \frac{0,024131}{0,024131 + 0,000728}$$

$$= 0,9707$$

$$P(X|C_1)P(C_1) = \frac{Jumlah(class=1)}{(Jumlah(class=0) + Jumlah(class=1))}$$

$$= \frac{0,000728}{0,024131 + 0,000728}$$

$$= 0,0292$$

Keputusan = “**negatif**”.

Tabel 4.16 Probabilitas Klasifikasi NBC pada Bukalapak.

Testing Komentar	Probabilitas Negatif	Probabilitas Positif	Keputusan
1	0,9707	0,0292	Negatif
2	0,7250	0,2250	Negatif
3	0,6350	0,3650	Negatif
4	0,1750	0,8250	Positif
5	0,8450	0,1550	Negatif
6	0,2850	0,7150	Positif
7	0,3450	0,6550	Positif
8	0,1250	0,8750	Positif
9	0,3750	0,6250	Positif
.	.	.	.
.	.	.	.
.	.	.	.
298	0,2250	0,7750	Positif

Berdasarkan Tabel 4.16 dapat diketahui probabilitas klasifikasi dengan metode *Naïve Bayes Classifier* (NBC) pada *online Shop* Bukalapak. Jika probabilitas sentimen negatif lebih tinggi jika dibandingkan dengan probabilitas sentimen positif, maka akan memberikan hasil keputusan negatif begitupula sebaliknya. Pada data komentar yang pertama, probabilitas sentimen negatif adalah sebesar 0,9707, sedangkan probabilitas sentimen positif adalah sebesar

0,0292. Keputusan yang dapat diambil adalah kalimat pertama masuk kedalam sentimen negatif, dikarenakan probabilitas sentimen negatif lebih besar jika dibandingkan dengan probabilitas sentimen positif.

C. Klasifikasi *Naive Bayes* pada Online Shop Blibli.com.

Untuk menentukan suatu komentar masuk ke dalam klas negatif (0) atau kelas positif (1) dapat dilihat dari probabilitas berikut $P(X|C_i)P(C_i)$. Model *Naive Bayes Classifier* (NBC) yang digunakan dalam penelitian ini adalah sebagai berikut.

Komentar pertama setelah proses *stemming* : “belanja, blibli, emang, mudah, banget, tambah, produk, jual, cukup, murah, tersedia, pilihan, pembayaran, tempat, bantu, buat, pemalu, takut, barang, tidak, terima, semua, pelayanan, bagus”.

Tabel 4.17 NBC Blibli.com

Kata	Positif	Negatif
Belanja	21	1
Blibli	55	1
Mudah	25	2
Banget	3	1
Produk	7	1
Jual	3	1
Murah	12	2
Tersedia	2	1
Pilihan	4	1
Pembayaran	16	1
Pemalu	1	2
Takut	2	1
Barang	34	1

Terima	1	1
Pelayanan	7	1
Bagus	10	2
Total	203	20

$$\text{Class positif (1)} = \frac{21}{22} \times \frac{55}{56} \times \frac{25}{27} \times \frac{3}{4} \times \frac{7}{8} \times \frac{3}{4} \times \frac{12}{14} \times$$

$$\frac{2}{3} \times \frac{4}{5} \times \frac{16}{17} \times \frac{1}{3} \times \frac{2}{3} \times \frac{34}{35} \times \frac{1}{2} \times \frac{7}{8} \times \frac{10}{12} \times \frac{66}{75}$$

$$= \frac{13300277760000}{919315198771200} \times \frac{66}{75}$$

$$= \frac{877818332160000}{68948639907840000}$$

$$= 0,12731$$

$$\text{Class negatif (0)} = \frac{1}{22} \times \frac{1}{56} \times \frac{2}{27} \times \frac{1}{4} \times \frac{1}{8} \times \frac{1}{4} \times \frac{2}{14} \times$$

$$\frac{1}{3} \times \frac{1}{5} \times \frac{1}{17} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{35} \times \frac{1}{2} \times \frac{1}{8} \times \frac{2}{12} \times \frac{9}{75}$$

$$= \frac{16}{919315198771200} \times \frac{9}{75}$$

$$= \frac{144}{68948639907840000}$$

$$= 0,00208$$

Naïve bayes classifier (Normalisasi) :

$$P(X|C_0)P(C_0) = \frac{\text{Jumlah (class=0)}}{(\text{Jumlah (class=0)} + \text{Jumlah (class=1)})}$$

$$= \frac{0,00208}{0,00208 + 0,12731}$$

$$= 0,0160$$

$$P(X|C_1)P(C_1) = \frac{\text{Jumlah (class=1)}}{(\text{Jumlah (class=0)} + \text{Jumlah (class=1)})}$$

$$= \frac{0,12731}{0,00208 + 0,12731}$$

$$= 0,9839$$

Keputusan : “**positif**”.

Tabel 4.18 Probabilitas Klasifikasi NBC pada Blibli.com.

Testing Komentar	Probabilitas Negatif	Probabilitas Positif	Keputusan
1	0,0160	0,9839	Positif
2	0,000	1,000	Positif
3	0,2150	0,7850	Positif
4	0,3350	0,6650	Positif
5	0,1500	0,8500	Positif
6	0,2350	0,7650	Positif
7	0,3050	0,6950	Positif
8	0,1250	0,8750	Positif
9	0,000	1,000	Positif
.	.	.	.
.	.	.	.
.	.	.	.
75	0,3750	0,6250	Positif

Berdasarkan Tabel 4.18 dapat diketahui probabilitas klasifikasi dengan metode *Naïve Bayes Classifier* (NBC) pada *online Shop* Blibli.com. Jika probabilitas sentimen negatif lebih tinggi jika dibandingkan dengan probabilitas sentimen positif, maka akan memberikan hasil keputusan negatif begitupula sebaliknya. Pada data komentar yang pertama, probabilitas sentimen negatif adalah sebesar 0,0160, sedangkan probabilitas sentimen positif adalah sebesar 0,9839. Keputusan yang dapat diambil adalah kalimat pertama masuk kedalam sentimen positif, dikarenakan probabilitas sentimen positif lebih besar jika dibandingkan dengan probabilitas sentimen negatif.

Tabel 4.19 Hasil Perhitungan NBC Pada situs belanja.

No.	Situs belanja	Perhitungan <i>Naive Bayes</i>	
		Positif	Negatif
1.	Lazada	95%	5%
2.	Bukalapak	92%	8%
3.	Blibli	88%	12%

4.4 Uji Validitas Data

Pengujian pada penelitian ini menggunakan *tools RapidMiner* untuk menghasilkan suatu output klasifikasi dan keakurasian data. *10-fold cross validation* digunakan dalam penelitian ini untuk menguji keakuratan data.

A. *10-fold cross validation* pada Online Shop Lazada.

Langkah selanjutnya adalah membandingkan hasil tingkat akurasi dengan membagi data training dan testing berdasarkan metode *10-fold cross validation*. Pada *10-fold cross validation*, data dibagi menjadi 10 *fold* kemudian data dibagi menjadi *training testing* dengan perbandingan 90:10 dengan metode sampling stratifikasi. Berikut adalah hasil ketepatan klasifikasi dengan menggunakan *10-fold cross validation* pada online Shop Lazada.

Tabel 4.20 Ketepatan Klasifikasi metode NBC dengan Menggunakan *10-Fold Cross Validation* pada Lazada.

Fold Ke-	Kriteria Penilaian
	Akurasi (Rapidminer)
1	98%
2	97%
3	95%
4	95%
5	98%
6	99%
7	97%

Fold Ke-	Kriteria Penilaian
	Akurasi (Rapidminer)
8	95%
9	95%
10	96%
Rata-Rata	96,5%

B. 10-fold cross validation pada Online Shop Bukalapak.

Langkah selanjutnya adalah membandingkan hasil tingkat akurasi dengan membagi data *training* dan *testing* berdasarkan metode *10-fold cross validation*.

Pada *10-fold cross validation*, data dibagi menjadi 10 *fold* kemudian data dibagi menjadi *training testing* dengan perbandingan 90:10 dengan metode sampling stratifikasi. Berikut adalah hasil ketepatan klasifikasi dengan menggunakan *10-fold cross validation* pada *online Shop Bukalapak*.

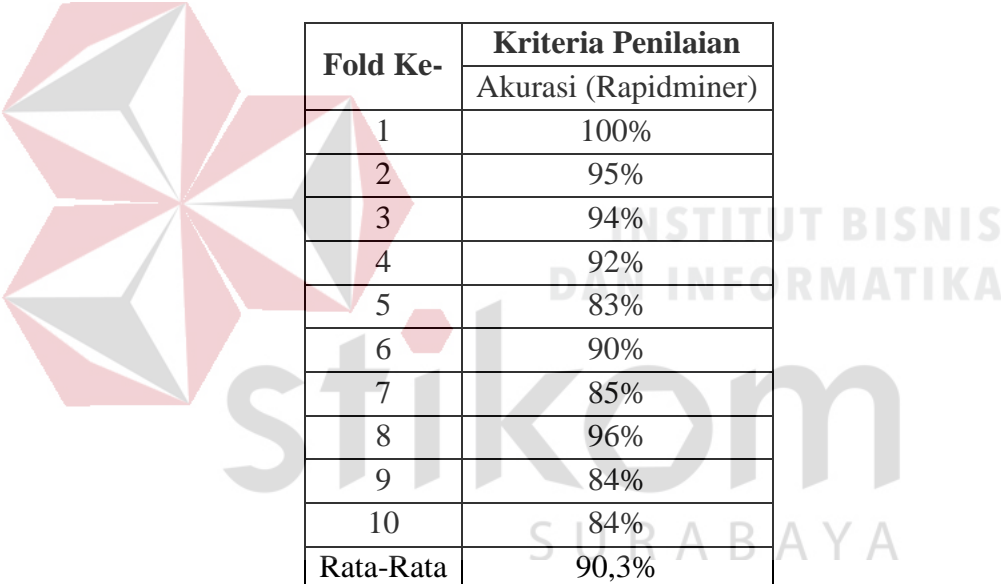
Tabel 4.21 Ketepatan Klasifikasi metode NBC dengan Menggunakan *10-Fold Cross Validation* pada Bukalapak.

Fold Ke-	Kriteria Penilaian
	Akurasi (Rapidminer)
1	92%
2	93%
3	99%
4	95%
5	98%
6	92%
7	91%
8	95%
9	92%
10	92%
Rata-Rata	93,9%

C. 10-fold cross validation pada Online Shop Blibli.com.

Langkah selanjutnya adalah membandingkan hasil tingkat akurasi dengan membagi data training dan testing berdasarkan metode *10-fold cross validation*. Pada *10-fold cross validation*, data dibagi menjadi 10 *fold* kemudian data dibagi menjadi *training testing* dengan perbandingan 90:10 dengan metode sampling stratifikasi. Berikut adalah hasil ketepatan klasifikasi dengan menggunakan *10-fold cross validation* pada *online Shop Blibli.com*.

Tabel 4.22 Ketepatan Klasifikasi metode NBC dengan Menggunakan *10-Fold Cross Validation* pada Blibli.com.



Fold Ke-	Kriteria Penilaian
	Akurasi (Rapidminer)
1	100%
2	95%
3	94%
4	92%
5	83%
6	90%
7	85%
8	96%
9	84%
10	84%
Rata-Rata	90,3%

4.5 Rating secara umum

- I. Berikut adalah paparan rating secara umum bila dilihat berdasarkan komentar pengguna website priceprice.com selama kurun waktu 2 tahun terakhir :

Tabel 4.23 Rating secara umum komentar priceprice.

Rank	Situs	Comment Count (cc)	Rating
1	Lazada	673	1
2	Bukalapak	298	2
3	Blibli.com	75	3

Tabel 4.23 menjelaskan bahwa secara *Comment Count* (cc) Lazada menempati peringkat 1 untuk *online shop* yang paling banyak disebut selama kurun waktu 2 tahun terakhir.

- II. Berikut adalah rating untuk komentar positif paling banyak berdasarkan komentar pengguna pada website priceprice.com setelah proses uji validitas data :

Tabel 4.24 Rating secara umum berdasarkan komentar positif.

Rank	Situs	Positive Comments (pc)	Rating
1	Lazada	95%	1
2	Bukalapak	94%	2
3	Blibli.com	88%	3

Tabel 4.24 menjelaskan bahwa secara *Positive Comments* (pc) Lazada menempati peringkat 1 untuk *online shop* yang paling banyak mendapatkan komentar positif oleh pengguna website priceprice.

- III. Berikut adalah rating untuk situs belanja online yang paling populer dan paling banyak disebut pada kolom komentar *website priceprice.com* :

Tabel 4.25 Rating secara umum berdasarkan paling populer.

Rank	Situs	Popular Online Shop (pos)	Rating
1	Lazada	95,9%	1
2	Bukalapak	93,7%	2
3	Blibli.com	89,3%	3

Tabel 4.25 menjelaskan bahwa secara *Popular Online Shop* (pos) Lazada menempati peringkat 1 untuk *online shop* yang paling populer dibandingkan dengan bukalapak dan Blibli.com menurut komentar pengguna *website priceprice.com*.

- IV. Berikut adalah rating untuk situs belanja online yang paling aman dan paling banyak disebut pada kolom komentar *website priceprice.com* :

Tabel 4.26 Rating secara umum berdasarkan paling aman.

Rank	Situs	Tingkat Keamanan (TK)	Rating
1	Lazada	79%	3
2	Bukalapak	93%	1
3	Blibli.com	92%	2

Tabel 4.26 menjelaskan bahwa secara Tingkat Keamanan (TK) Bukalapak menempati peringkat 1 untuk *online shop* yang paling aman menurut komentar pengguna *website priceprice*.

- V. Berikut adalah rating untuk situs belanja online yang paling memuaskan dan paling banyak disebut pada kolom komentar *website priceprice.com* :

Tabel 4.27 Rating secara umum berdasarkan paling memuaskan.

Rank	Situs	Kepuasan Pelanggan (KP)	Rating
1	Lazada	91%	1
2	Bukalapak	69%	3
3	Blibli.com	58%	2

Tabel 4.27 menjelaskan bahwa secara Kepuasan Pelanggan (KP) Lazada menempati peringkat 1 untuk *online shop* yang paling memuaskan berdasarkan komentar pengguna website priceprice.

Berdasarkan pemeringkatan yang dilakukan pada tahap ini, didapatkan bahwa *online shop* Lazada memimpin sebagai *online shop* paling populer diantara Bukalapak dan Blibli.com menurut data dari *platform website* priceprice.com dan metode perhitungan *naive bayes*. Berikut tampilan pemeringkatan, dapat dilihat pada tabel 4.28

Tabel 4.28 Pemeringkatan Keseluruhan.

No.	Kategori	Lazada	Bukalapak	Blibli.com
1	<i>Comment Count (cc)</i>	1	2	3
2	<i>Positive Comments (pc)</i>	1	2	3
3	<i>Popular Online Shop (pos)</i>	1	2	3
4	Tingkat Keamanan (TK)	3	1	2
5	Kepuasan Pelanggan (KP)	1	2	3

4.6 Visualisasi data dengan Wordcloud

Pendapat dari konsumen belanja online di Lazada, Bukalapak dan Blibli.com dapat divisualisasikan dalam bentuk *Wordcloud*. *Wordcloud* merupakan representasi grafis dari sebuah dokumen yang dilakukan dengan plotting kata-kata yang sering muncul pada sebuah dokumen pada ruang dua dimensi. Frekuensi dari kata yang sering muncul ditunjukkan melalui ukuran huruf kata tersebut. Semakin besar ukuran kata menunjukkan semakin besar frekuensi kata tersebut muncul dalam dokumen. Berikut adalah tampilan visualisasi data dengan *Wordcloud* berdasarkan hasil uji validitas.

A. Lazada



Gambar 4.11 Visualisasi Wordcloud Lazada.

Berdasarkan Gambar 4.11 dapat diketahui bahwa tiga kata yang sering muncul pada sentimen positif akun Lazada adalah Lazada, barang, dan belanja. Peningkatan pelayanan dan kecepatan kepada pelanggan dari *online shop* Lazada perlu dipertahankan dan diberi apresiasi. Selain respon positif, respon negatif juga perlu diketahui oleh Lazada untuk meningkatkan pelayanan.

B. Bukalapak

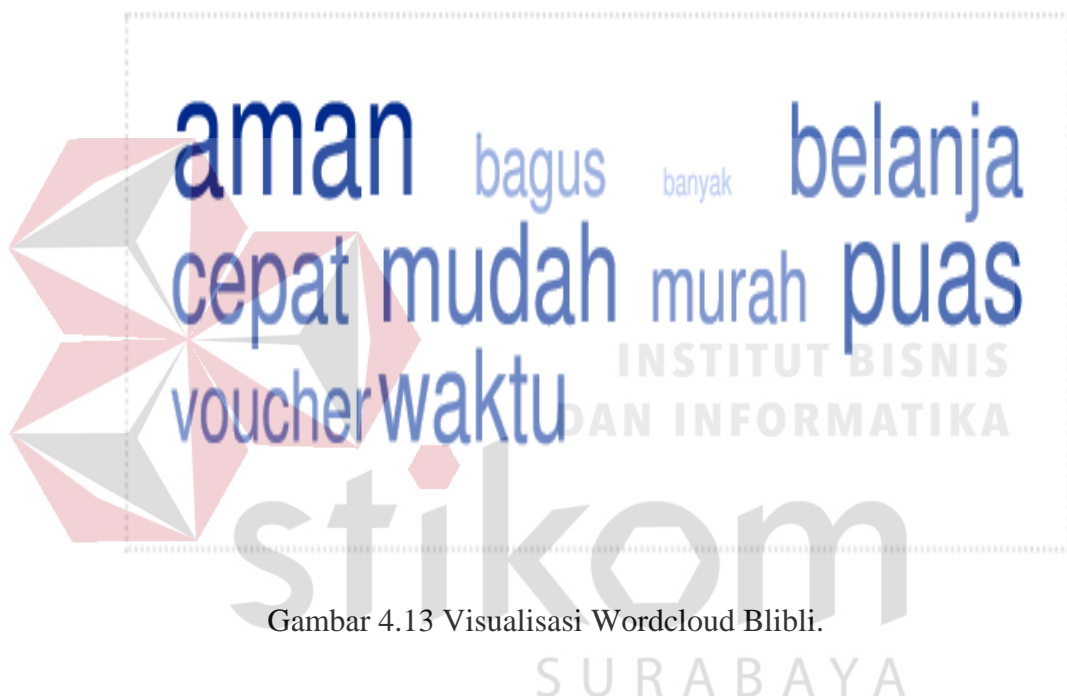


Gambar 4.12 Visualisasi Wordcloud Bukalapak.

Berdasarkan Gambar 4.12 dapat diketahui bahwa tiga kata yang sering muncul pada sentimen positif akun Bukalapak adalah bukalahpak,

belanja, dan barang. Peningkatan pelayanan tentang bantuan dan memberikan informasi kepada masyarakat dari *online shop* Bukalapak perlu dipertahankan dan diberi apresiasi. Selain respon positif, respon negatif juga perlu diketahui oleh Bukalapak untuk meningkatkan keamanan pelanggan dalam bertransaksi.

C. Blibli.com



Gambar 4.13 Visualisasi Wordcloud Blibli.

Berdasarkan Gambar 4.13 dapat diketahui bahwa tiga kata yang sering muncul pada sentimen positif akun Blibli.com adalah blibli, barang, dan aman. Keamanan dalam bertransaksi Blibli.com perlu dipertahankan dan diberi apresiasi. Selain respon positif, respon negatif juga perlu diketahui oleh Blibli.com untuk meningkatkan kepuasan pelanggan dalam berbelanja.

BAB V

PENUTUP

Berdasarkan hasil analisis yang telah diperoleh, kesimpulan dan saran yang didapatkan dari penelitian ini adalah sebagai berikut.

5.1 Kesimpulan

Berdasarkan hasil analisis yang telah diperoleh, kesimpulan yang didapatkan dari penelitian ini adalah sebagai berikut.

1. Konsumen yang memberikan tanggapan, pendapat, kritik, saran dan masalah *complain* lebih banyak ditujukan pada *online shop* Blibli.com dari pada Bukalapak maupun Lazada. Jumlah sentimen data pada ketiga *online shop* bersifat *imbalanced*, sehingga ketepatan klasifikasi perlu dilihat berdasarkan uji validitas data.
2. Berdasarkan visualisasi *wordcloud*, penulis mengambil 3 sentimen positif teratas yang banyak dibicarakan oleh konsumen pada *online shop* :
 - a) Lazada : baik, murah, cepat.
 - b) Bukalapak : terima kasih, barang, belanja.
 - c) Blibli.com : aman, puas, mudah.
3. *Online Shop* yang paling populer menurut penelitian ini adalah Lazada dengan rata-rata penghitungan manual, sentimen positif 95%, disusul dengan Bukalapak dengan rata-rata sentimen positif 94%, kemudian di posisi paling akhir adalah Blibli.com dengan rata-rata sentimen 88%.
4. Pada penelitian ini Lazada merupakan *online shop* yang paling populer dibandingkan dengan Bukalapak dan Blibli.com. Berdasarkan perhitungan

pemeringkatan popularitasnya didapatkan untuk *online shop* Lazada memiliki tingkat popularitas 95,9%, Bukalapak 93,7% dan Blibli.com 89,3%.

5. Dengan uji validitas metode NBC menggunakan *10-fold cross validation* didapatkan rata-rata untuk Lazada 96,5%, Bukalapak 93,9% dan Blibli.com 90,3%. Artinya, metode NBC memiliki tingkat akurasi tinggi dalam klasifikasi sentimen analisis.
6. Berdasarkan pemeringkatan yang dilakukan, didapatkan bahwa *online shop* Lazada memimpin sebagai *online shop* paling populer diantara Bukalapak dan Blibli.com menurut data dari *platform website priceprice.com* dan metode perhitungan *naive bayes*. Dapat dilihat pada tabel 4.28

5.2 Saran

Berdasarkan hasil analisis yang telah dijelaskan, maka saran yang dapat diberikan pada penelitian yang akan datang adalah sebagai berikut.

1. Pada penelitian klasifikasi teks, membutuhkan data sentimen yang cukup besar untuk mengurangi resiko dari imbalanced data. Pengklasifikasian sentimen awal sebaiknya dilakukan dengan banyak orang sehingga dapat mengurangi obyektifitas pada hasil sentimen.
2. Pada proses *stemming*, perlu merubah kata tidak baku yang digunakan untuk bahasa sehari-hari dikarenakan pada data komentar banyak menggunakan bahasa sehari-hari. Apabila ditemukan sedikit kata tidak baku maka dapat merubah data secara manual saja, jika sebaliknya maka harus menggunakan *tools* khusus.

3. Penulis sangat menyarankan untuk menggunakan metode NBC pada penelitian yang akan datang dikarenakan metode ini lebih baik dari pada metode lainnya, akurasi yang dihasilkan pada metode NBC sangat besar apabila dibandingkan dengan metode lain yang serupa.

Kemudian, saran untuk *online shop* yang dibahas dalam penelitian ini adalah :

1. Lazada : Pertahankan pencapaian hingga kini. Karena Lazada menempati peringkat pertama untuk online shop yang paling populer seperti dibahas dalam penelitian ini. Namun, perlu diperhatikan untuk keamanan dalam transaksinya.
2. Bukalapak : Bukalapak menempati peringkat kedua dalam penelitian ini. Pertahankan dalam menjaga kepuasan pelanggan karena pada hal ini Bukalapak menempati peringkat kedua. Perlu diperhatikan juga untuk keamanan dalam transaksi.
3. Blibli.com : Sejauh ini Blibli.com menempati peringkat 1 untuk *online shop* yang paling aman dalam transaksinya. Namun, masih jauh dibelakang dibanding dengan Lazada dan Bukalapak untuk tingkat popularitas, dan masih banyak ulasan-ulasan negatif mengenai Blibli.com. Diharapkan ada suatu pembenahan untuk meningkatkan kualitas layanan dan lebih mengutamakan kepuasan pelanggan.

DAFTAR PUSTAKA

Alpaydin, Ethem. (2014). *Introduction to Machine Learning, Third Edition*.

London: The MIT Press.

E-marketer. (2014). *Pengguna Internet Indonesia Nomor Enam Dunia*. Tersedia di

<http://tekno.kompas.com>. (diakses pada 24/07/2018).

Ervinzein. (2016). *E-commerce-bibli-com*. Tersedia di

<https://Ervinzeintugas.wordpress.com>. (diakses pada 24/07/2018).

Gamallo, Pablo, & Fernandez-Lanza. (2013). TASS: A Naive-Bayes strategy for sentiment analysis on Spanish tweets. In In Workshop on Sentiment Analysis at SEPLN (TASS2013) (pp. 126–132). Madrid, Spain.

L. Whitten, Jeffrey, & D. Bentley, Lonnie .(2007). *System Analysis & Design Methods Seventh Edition*. New York, USA : McGraw-Hill

Liu, *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publisher, (2012).

Makinudin dan Hadiyanto (2006). *Analisis Sosial*. Bandung : Yayasan Akatiga.

McLeod, Pearson (2008). *Sistem Informasi Manajemen*. Salemba. Jakarta.

Pang, B. dan Lee, L., 2008, *Opinion Mining And Sentiment Analysis*, Foundations and Trends in Information Retrieval, 2, 1-2, 1-135.

Rizqiyah, Putri. (2018). *Analisa Sentimen Menggunakan Naïve Bayes Untuk Mengetahui Presentase Komentar Pada Aplikasi Go-Jek*. Program Studi Informatika, Institut Teknologi Telkom Purwokerto

Sanianto. (2014). *Profil Perusahaan Besar Lazada*. Tersedia di <http://irvanarifsanianto.blogspot.com/2014/11/perusahaan-lazada.html>

Saraswati, Ni Wayan Sumartini (2011). *Text Mining Dengan Metode Naïve Bayes Classifier Dan Support Vector Machines Untuk Sentiment Analysis*. Tesis [Online]. Tersedia di http://www.pps.unud.ac.id/thesis/pdf_thesis/unud-209-236721286-tesis.pdf. (diunduh pada 13/09/2015).

Sinaga, Ardiles dan Murnawan. 2017. *Decision Support System Model Analysis for Proposed Activities on Development Planning Forum In District Level*. IEEE. doi 10.1109_CITSM.2016.7577522.

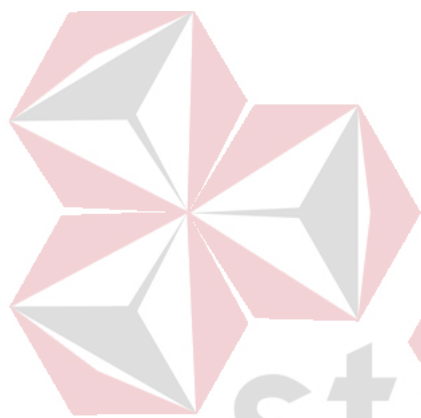
Silviosilviando. (2016). *Analisis web e-commerce bukalapak*. Tersedia di <https://silviosilviando.wordpress.com>. (diakses pada 23/5/2018).

Triawati, C (2009). *Metode Pembobotan Statistical Concept Based untuk Klastering dan Kategorisasi Dokumen Berbahasa Indonesia*. Institut Teknologi Telkom Bandung.

Vinodhini, G. And RM. Chandrasekaran (2012). "Sentiment Analysis and Opinion Mining: A Survey". International Journal of Advanced Research in Computer Science and Software Engineering, Vol.2, No.6, pp.283-292.

Wong, Jony. (2010). *Internet Marketing for Beginners*. PT Elex Media Komputindo. Jakarta.

Xhemali, D., Hinde, C.J. & Stone, R.G. 2009. *Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages*. International Journal of Computer Science Issues 4(1): 16-23. (Online) <http://ijcsi.org/papers/4-1-16-23.pdf> (16 Mei 2013)



INSTITUT BISNIS
DAN INFORMATIKA

stikom
SURABAYA